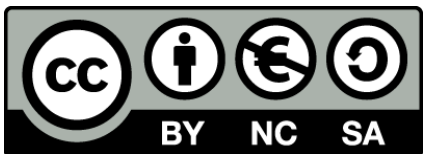
UNIVERSITAT DE BARCELONA

Structural and functional analysis of Zika Virus NS5 protein

Víctor Manuel Ruiz Arroyo



TESIS DOCTORAL

Structural and functional analysis of Zika Virus NS5 protein

Victor Manuel Ruiz Arroyo

UNIVERSITAT DE
BARCELONA

Año 2020

Structural and functional analysis of Zika Virus NS5 protein

PROGRAMA DE DOCTORADO DE BIOTECNOLOGIA

FACULTAD DE FARMACIA Y CIENCIAS DE LA ALIMENTACION

Doctorando: Victor Manuel Ruiz Arroyo

Directora: Nuria Verdaguer Massana

Director: Diego Sebastian Ferrero

Tutora: Josefa Badia Palacin

Agradecimientos:

Me gustaría agradecer a todas las personas que se han cruzado conmigo durante estos cuatro años, para bien y para mal, porque al final, de todo se aprende.

A Nuria por haberme dado la oportunidad de hacer la tesis en su laboratorio cuando ya todo parecía perdido y ninguna puerta se abría.

A Diego por la inmensidad de conocimiento que me ha transmitido y por haber aguantado durante estos cuatro años todas mis meteduras de pata.

A mis compañeros de Cri6: A Pablo por enseñarme cryo-EM y a aguantar como Diego mis meteduras de pata y por sus momentos musicales. A María por todos los buenos ratos e idas de olla continuas, a Omar y su sonrisa continua, a Cristina por las conversaciones diarias en el laboratorio de temas del barrio, a Mónica por los buenos ratos en las “calçotadas” y en los “diables”.

A Aleix por ser compañero “a la par” desde el comienzo de esta aventura, por tantos viajes al sincrotrón, subidas en bici por Collserola con su correspondiente “paraeta en Can Coll”. Al resto de “chiques” de Cri7 por su petardeo continuo.

A Ira y Elena, son tantísimos los buenos momentos por los que os estoy agradecido que nunca acabaría, por los gabinetes de crisis tan resolutivos, por los fines de semana 3/3, los conciertos non-stop, por la despedida tan genial que organizasteis, por los 28 de Junio pasados y el que nos ha faltado este año, muchísimas gracias por estar ahí siempre.

A Michela, la italiana loca que tanto me alegro los primeros años, importadora del piti a las 12 y de las fiestas italianas en la terraza del ático.

A todos los compañeros que dejé en Canadá, a Yao por todas las horas que pasó conmigo en el microscopio, al pinche Javi por las noches de tapitas de

tequila, a Lau por iniciarme al gimnasio, a Srhreya por su inmensa simpatía y positividad. A Alba por darme la grandísima oportunidad de estar en su laboratorio. Hicisteis de mi estancia una gran experiencia.
A Loïc porque me recibió con una cerveza en la mano en plena tormenta de nieve. Porque más que un compañero de piso fue un amigo, gracias por las conversaciones hasta altas horas de la mañana que me hicieron hablar ingles y a recordar el francés. Por los momentos musicales, por el hockey, por las quishes y por acercarme a la cultura local. “See you soon man”.

A toda mi familia, porque a pesar de que estos cuatro años hemos estado separados, hemos estado mas unidos que nunca. A mi hermana por su apoyo incondicional durante todo este camino. A mis padres por estar ahí siempre y haberme apoyado y desvivirse por mi hasta en los peores momentos. A Jose porque los hermanos a veces se eligen.

A Inma, Carol y Loles, que sin ellas desde un principio esto nunca hubiese sido posible. Gracias por enseñarme y alentarme en mis etapas tempranas como investigador, por confiar en mi y ayudarme en todo lo necesario.

A Joan, porque cuando este libro se cierre se abrirá uno nuevo en blanco que está aun por escribir.

Table of Contents

Abstract

Zika virus (ZIKV) belongs to the *Flaviviridae* family and constitute an important public health concern since ZIKV infection produced devastating effects in new born infants. Flaviviruses present a positive sense single stranded RNA genome flanked by highly structured untranslated regions (UTR) carrying one open reading frame that codifies for three structural proteins (C, prM, E) and five non-structural proteins (NS1-5). At the most C-terminal end, NS5 protein carries a RNA dependent RNA polymerase (RdRP) and a methyl transferase domain (MTase) for genome copying and 5' capping activities of the newly synthesized RNA, respectively. Given the crucial role of this enzyme for viral replication, NS5 constitutes an attractive antiviral target to inhibit viral replication. In this study, we determined the structure of the ZIKV NS5 protein using X-Ray crystallography combined with several structural biology approaches to characterize the supramolecular arrangement of the ZIKV NS5 protein. We identified the monomer-monomer and dimer-diner interactions to form fibril-like structures, and evaluated the role of oligomer formation, using *in-vitro* polymerization assays. We also evaluated the *in-vivo* effect of NS5-oligomerisation in chicken embryos, stablishing a connection between this protein and microcephaly.

One of the most important RNA structures present at the 5'UTR of flavivirus genomes is the 5SLA. This structure was identified previously to bind the NS5 protein, acting as a promoter and being essential for viral replication. We assayed and optimized the NS5-5SLA complex stability using biophysical and biochemical techniques and determined the structure of the complex by single particle cryo-EM. Comparisons between the NS5-5SLA complex and the NS5 crystallographic structure revealed for the first time in flavivirus, important conformational changes in the NS5 RdRP. We identified the residues involved in complex formation and characterized the effect of this binding on NS5 polymerization, shedding new light on the understanding of replication mechanisms in flaviviruses.

Resumen:

El virus Zika (ZIKV) pertenece a la familia *Flaviviridae* y constituye una amenaza para la salud pública, especialmente debido a las malformaciones provocadas en neonatos. Los flavivirus presentan un genoma RNA de simple cadena con polaridad positiva, flanqueado por regiones no traducidas (UTR) que presentan una elevada estructura secundaria, seguido de una región codificante para una única poliproteína que por proteólisis dará lugar a tres proteínas estructurales (C, prM, E) y cinco proteínas no estructurales (NS1-5). En el extremo C-terminal se encuentra la proteína NS5 que presenta actividad ARN polimerasa dependiente de ARN (RdRP) y un dominio metil-transferasa (MTase) para copiar el genoma y añadir una caperuza al extremo 5' del nuevo ARN sintetizado, respectivamente. Dado el papel crucial de este enzima en la replicación viral, la proteína NS5 constituye una diana antiviral muy atractiva para inhibir la replicación del virus. En este estudio, determinamos la estructura de la proteína NS5 de ZIKV, usando cristalografía de Rayos-X combinada con diferentes técnicas biofísicas para caracterizar la organización supramolecular de la proteína. Identificamos las interacciones monómero-monómero y dímero-dímero para caracterizar las estructuras fibrilares de la proteína y evaluamos los efectos de la dimerización en la actividad polimerasa *in-vitro.* También evaluamos los efectos de la oligomerización de NS5 *in-vivo* en embriones de pollo, estableciendo una conexión entre esta proteína y la aparición de microcefalia en fetos infectados.

Una de las estructuras de ARN más importantes presentes en el 5'UTR del genoma de los flavivirus es el 5SLA. Previamente se describió que esta estructura se unía a NS5 y actuaba como un promotor, siendo además esencial para la replicación viral. Medimos y optimizamos la estabilidad del complejo NS5-5SLA mediante técnicas biofísicas y bioquímicas y determinamos la estructura del complejo mediante cryo-EM. Las comparaciones entre la estructura cristalográfica y cryo-EM de NS5 revelaron, por primera vez en flavivirus, cambios conformacionales importantes en el dominio RdRP. Identificamos los residuos implicados en la formación del complejo y caracterizamos el efecto de la unión de NS5 a 5SLA sobre su actividad polimerasa. Estos resultados arrojan nueva luz para entender los mecanismos de replicación en los flavivirus.

1 Introduction

1.1 Viruses

Viruses are the smallest self-replicating organisms, with sizes ranging from the tiniest Mouse Minute Virus that is only 26 nm in diameter to the largest Giant Mimivirus with a diameter of 750 nm. However, the most basic form consists of a nucleic acid genome enclosed in a protein coat and, for some viruses, surrounded by a membranous envelope. Viruses are obligate parasites that manipulate the host machinery to their own purposes. Despite its simplicity, viruses have evolved to infect virtually every living organism. This diversity has constantly challenged to scientists to change their research and classification methods to better understand the behavior of these organisms. Viruses can be classified attending phenotypic characteristics, such as morphology, nucleic acid type, mode of replication, host organisms or type of disease they produce. The Baltimore classification system is an easy and common way to place a virus in one of the seven groups based on their manner of mRNA synthesis: i) double-stranded DNA (dsDNA), ii) single stranded DNA (ssDNA), iii) double-stranded RNA (dsRNA), iv) positive single-stranded RNA (+ssRNA), v) negative single-stranded RNA (-ssRNA), vi) single stranded RNA with a DNA intermediate (ssRNA-RT) and vii) double stranded DNA with an RNA intermediate (dsDNA-RT) (Baltimore 1971).

+ssRNA viruses represents a large group of viruses which include most of the causative agents of important human diseases, such as poliovirus (PV), Hepatitis C virus (HCV), Zika virus (ZIKV) or the actual pandemic SARS Coronavirus-2.

1.2 The Flavivirus genus

The *Flaviviridae* family includes an extensive group of +ssRNA viruses that were first discovered one century ago by Walter Reed, who described that yellow fever could be transferred between patients via filtered serum and the infectious agent was transmitted by mosquitoes. The word *flavus* means “yellow” in Latin and the yellow fever, in turn, is named from its propensity to cause jaundice in sick

patients. Currently, this family contains four genera: *Flavivirus*, *Pestivirus*, *Hepacivirus* and *Pegivirus* (ICTV). The *Flaviviridae* family shares similarities in virion morphology, genome organization and replication strategies.

The *Flavivirus* genus are among the most important human pathogens and causes hundreds of millions of infections annually. They are spread globally and cause very wide spectrum of illness, from mild symptoms to severe and fatal hemorrhagic and neurological diseases. This genus comprises 53 viral species classified into three distinct groups with different transmission cycles: vector-borne flaviviruses (mosquito-borne and tick-borne) and viruses with not known arthropod vector. The largest and most important group is the mosquito-borne group with 27 species, followed by the tick-borne group with 12 species and the not known vector group which comprises 14 species (ICTV).

Mosquito-borne viruses contains the most relevant human pathogens such as Dengue Virus (DENV), Japanese Encephalitis Virus (JEV), West Nile Virus (WNV), Yellow Fever Virus (YFV) or Zika Virus (ZIKV). Mosquito-borne flavivirus transmission occurs after an arthropod feed on the infected vertebrate. An incubation period of 8-14 days after blood meal is necessary for the virus to replicate into the arthropod tissues. Virus transmission to another vertebrate occurs only after the salivary glands are infected and the virus is injected with the salivary fluid when the mosquito feed on other vertebrate host.

Flaviviruses have been historically an important public health concern for humans. Several epidemics of YFV and DENV have occurred regularly from the 17th to the 20th centuries (Mackenzie *et al.* 2004). While DENV epidemics occurred on most continents, YFV was limited to Africa, Europe and the Americas. Both viral diseases were transmitted by the African mosquito *Aedes aegypti* which extended to the rest of the continents as a consequence of the global commerce during the 18th and 19th centuries. Both YFV and DENV were effectively controlled during the 20th century throughout extensive mosquito eradication programs and the use of insecticides such as DDT, also an effective vaccination program for YFV controlled virus cases in West Africa. However,

DENV cases arised again during the 70's due to the termination of the eradication programs (Gubler 1998).

Nowadays, the increasing population specially in developed countries is the main driver of the environmental transformation, changing the agricultural, urbanization and people transport habits between others. These factors have contributed to a re-emergence of flavivirus-associated diseases allowing the dispersion of these pathogens causing major epidemics (Gubler 1998 and 2002). Some examples of this geographic expansion is the case of DENV, World Health Organization reported cases have increased nearly 10-fold in the last 40 years since 1970. JEV also has expanded from its endemic area towards Nepal, India, Papua New Guinea and Australia (Solomon et al 2003). WNV also expanded and jumped in 1999 to the Americas where it is now present (Artsob 2009). WNV is also another case of reemerging disease in the tropical areas of America and Africa (Duffy 2009, Gardner and Ryman 2010). Currently, flaviviruses represent one of the most important emergent tropical diseases, the number of associated cases raises up to tens of millions. However, there are no therapeutic options for the treatment of any flavivirus infection, and supportive care is the only treatment method that has shown some success.

1.2.1 Genome structure

Flavivirus genome is composed by a positive polarity (+) single-stranded RNA (ssRNA) molecule of ~11 kb that contains one single open reading frame (ORF) that acts as a mRNA, being directly transcribed by the host ribosomes. It harbors a conserved type I RNA cap (m7GpppA2'OmG) at the 5'-end and lacking of a poly(A) tail at its 3' end. Flanking both 5' and 3' ends, the genome contains two highly structured untranslated regions (UTRs), of ~100 nt at the 5' and ~400-700 nt at the 3' end (Rice *et al.* Science) (Figure 1). The 5'UTR harbors at 5' an Y-shaped RNA structure termed stem-loop A (5SLA) (Figure 2). Several studies detailed that 5SLA structure is composed by three helical regions (stem 1-3) and two single stranded regions: one top and side loop. The 5SLA structure is highly conserved in all flavivirus and is predicted to acts as promoter in the genome replication by direct recruiting the NS5 protein (Lodeiro *et al.* 2009). The second domain is separated by a uridine-rich sequence that acts as a spacer for proper

function of the two stem loops. The second domain extend from the 5'UTR into the C coding sequence, it contains the 5' upstream of AUG region (5'UAR) that folds into a second stem-loop structure (SLB) which contains essential sequences for long-range RNA-RNA interactions and genome replication (Alvarez *et al.* 2005) (Figure1). Another secondary structure RNA element can be found inside the C protein coding region, this feature is termed cHP and plays essential role in both translation and RNA synthesis (Clyde *et al.* 2008) and acts as enhancer of the recognition of the 5' C protein start codon (Clyde *et al.* 2006) (Figure1).

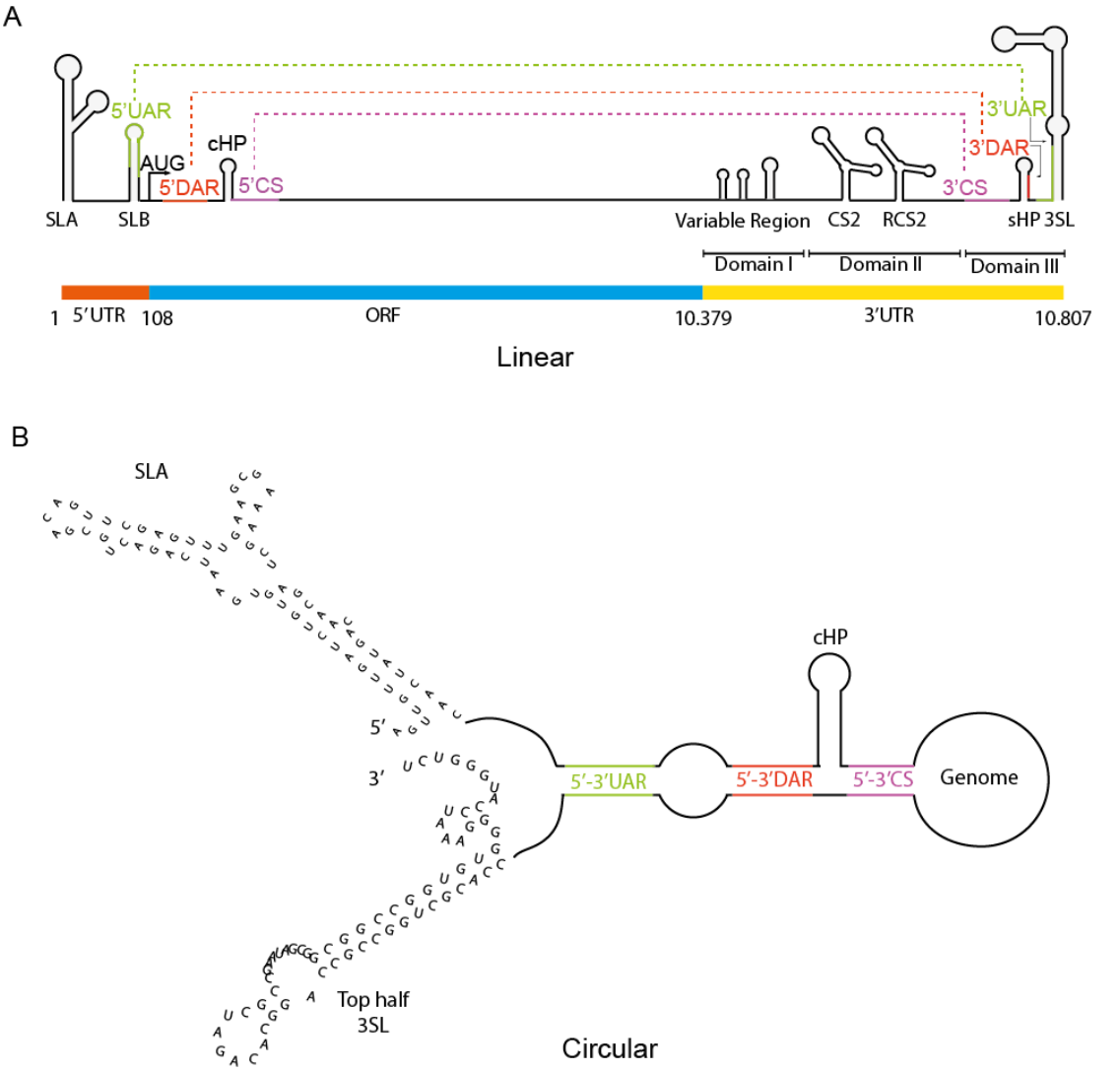


Figure 1: Schematic representation of the RNA secondary structures present in the flavivirus genome. **(A)** Linear conformation of Flavivirus genome. The 5' and 3' UTRs are colored in red and yellow respectively, the open reading frame (ORF) is colored in blue. Nucleotide position is indicated explicitly. Long range RNA-RNA interacting sequences: UAR (green), DAR (red) and CS (purple) are depicted in dashed lines. **(B)** Flavivirus circular genome conformation. The predicted folding for the 5SLA as well as the conformational changes that take place within the 3SL structure upon genome circularization are represented.

The 3'UTR region consist in approximately ~450nt and can be divided into three domains (Figure 1). Domain I is located just after the ORF stop codon and comprises the most variable region within the 3'UTR. Following this region, Domain II presents two tandem duplicated dumbbell sequences (DB) named CS2 and RCS2 (repeated CS2) (Figure 1) (Gritsun *et al.* 2006, Olsthoorn *et al.* 2001, Romero *et al.* 2006) that has been proposed to be involved in pseudoknot structure during viral translation and RNA synthesis (Manzano *et al.* 2011) and serve as enhancers of viral processes (Alvarez 2005, Bredenbeek 2003). Domain III is the most conserved region in the 3'UTR. It presents a terminal stem-loop structure (3'SL) that contains a short stem loop of 14 nt (sHP) followed by a large stem loop of 79 nt (Figure 1). The existence and the essential role for viral replication of the 3' stem loop structure in DENV and other Flavivirus has been has been confirmed and described extensively (Deng 1993, Hahn 1987, Takegami 1986, Rice 1985, Yu *et al.* 2005 and 2008).

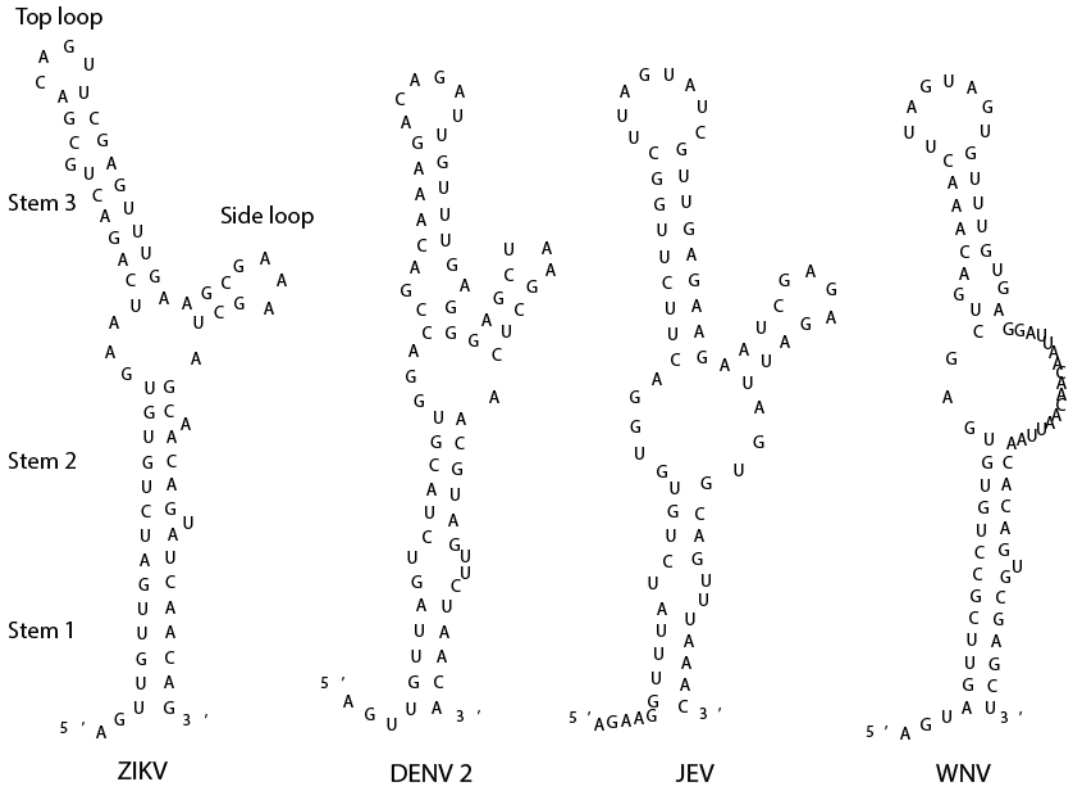


Figure 2: Domain description and predicted folding of the 5SLA structures in different flaviviruses (access codes: for ZIKV, MT507050; DENV 2, U87411; JEV, L48961 and WNV, M12294).

Flavivirus genome dynamics have been extensively studied using DENV 2 as model (Alvarez *et al.* 2005, Friebre *et al.* 2010, Polacek *et al.* 2009) and it is considered valid for the rest of flaviviruses. The viral genome presents three pairs of inverted complementary sequences at the genome ends termed 5'-3'CS, 5'-3'UAR and 5'3'DAR that are implicated in long RNA-RNA interactions that allow the circularization of the genome (Figure 1). The 5'CS sequence is 11 nt long and is located inside the ORF, overlapping the first residues of the capsid protein, it is complementary to the 3'CS sequence which places upstream of the 3'SL structure at the 3'UTR. The 5'UAR is located upstream of the translation initiator AUG in the 5'UTR and is complementary to the 3'UAR at the 3'UTR region, overlapping both the sHP and part of the 3' stem loop. The 5'DAR sequence is located downstream of the AUG and has been predicted to be complementary to the 3'DAR sequence at the 3'UTR.

The concerted hybridization of these three pair of sequences results in conformational changes in several RNA genomic structures. The 5'-3'CS sequence pair interacts forming a double stranded region which seems to be essential for afterward pair interactions. The 5'-3'UAR and 5'-3'DAR sequence pairs that overlap the SLB and sHP structures, form a duplex that loose these secondary structures and also open part of the 3' stem loop resulting in a unstructured 3' genome end (Figure 1). These long-range RNA interactions has been demonstrated biochemically (Fiebre *et al.* 2010, Polacek *et al.* 2008) and the genome circularization has been imaged using atomic force microscopy (AFM) (Filomatori *et al.* 2006), also, *in vivo* studies with infectious clones and replicon systems, have proven that genome cyclization is essential for flavivirus genome replication (Alvarez *et al.* 2005, Khromykh *et al.* 2001, Kofler *et al.* 2006, Lo *et al.* 2003).

1.2.2 Viral proteins

Flaviviruses enter the host cell by receptor-mediated endocytosis. The viral envelope (E) protein, a ~50 kDa membrane-anchored glycoprotein, coats the lipid envelope of mature virions and mediates both cell attachment and fusion. After fusion of viral and host cell membranes, the ribonucleoprotein is released

into the cytosol. Translation initiation of flavivirus genome is mediated by both canonical cap-dependent and non-canonical mechanism when the cap-dependent translation is suppressed (Edgil *et al.* 2006). Also, the cHP RNA structure, positioned at the starting AUG codon acts as enhancer to initiate translation (Clyde 2006).

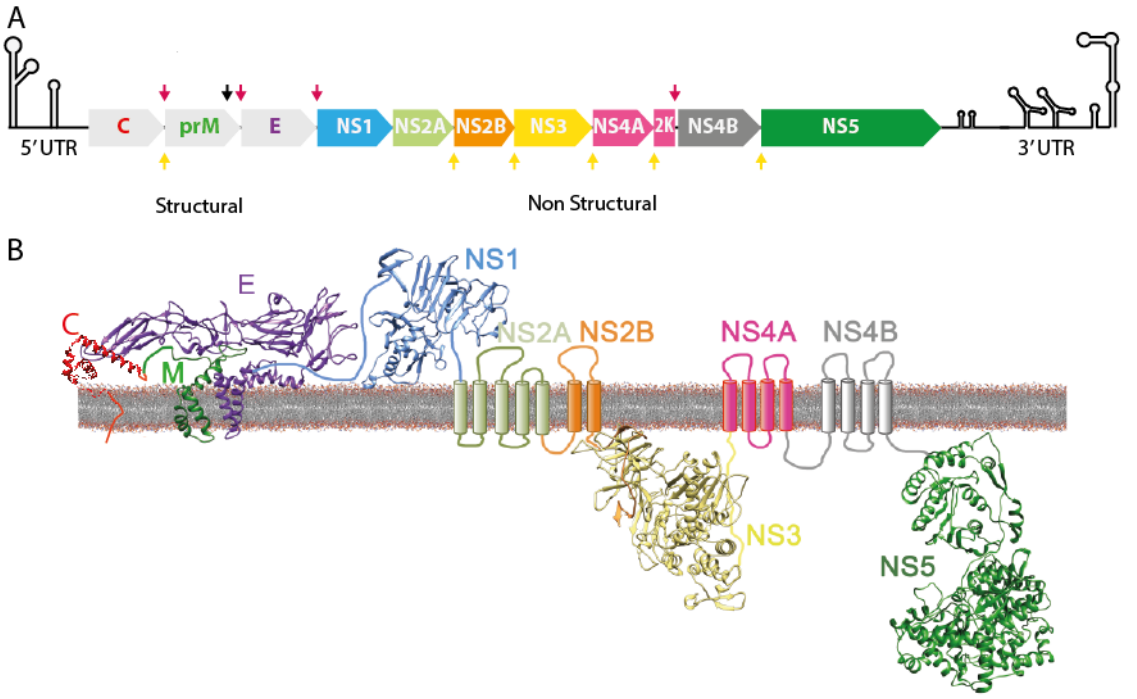


Figure 3: Schematic representation of the flavivirus ORF. **(A)** Fragments corresponding to individual proteins are depicted with arrowed boxes. The structural proteins: C, prM and E are shown in light grey, and the non-structural proteins are depicted in different colors: NS1 (blue), NS2A (light green), NS2B (orange), NS3 (yellow), NS4A (pink), NS4B (grey) and NS5 (green). NS3-mediated proteolytic cleavages are indicated with yellow arrows, while host peptidase and golgi mediated cleavages are colored in magenta and black respectively. **(B)** Illustration showing the membrane distribution of the structure models of ZIKV proteins colored using the same color code as in (A).

Genome translation takes place in the endoplasmic reticulum (ER), where the ORF is translated by host cell ribosomes, yielding one single polyprotein that is post-translationally cleaved by the host and viral protease into three structural proteins: capsid, pre-matrix/matrix and envelope (C, PrM/M, and E) and seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) (Figure 3). NS3 in complex with its cofactor NS2B is the catalytically active viral trypsin-like protease, and performs the proteolytic cuts between C-prM aimed by a cellular signalase, a cellular protease which cleaves cotranslationally (Johnson *et al.* 1999). The endogenous signalase mediates the cut between prM-E, E-NS1 and NS1-NS2. NS3 also mediates the cleavage in-cis of NS2A-NS2B and NS2B-

NS3 proteins while, for NS3-NS4A and NS4B-NS5 cleavage, NS3 acts in trans. Finally, NS4A and NS4B mature proteins are produced by a dual cleavage by NS3 and the cellular signalase, releasing the 2K peptide which inserts in the ER membrane (Rice *et al.* 1985) (Figure 3).

The virion consists in an enveloped particle of 50 nm in diameter that protects the viral genome. The outer layer of the virion is surrounded by 180 copies of the E glycoprotein. In addition to cell attachment and membrane fusion, the E protein is also involved in assembly and virus budding (Pierson *et al.* 2008). The precursor to membrane protein (prM) is a ~20 kDa protein that anchors to the viral membrane (Mukhopadhyay *et al.* 2005, Zhang *et al.* 2003) and helps in the folding of E protein (Roby *et al.* 2015). During virion formation, E and prM proteins form heterotrimeric spikes that come out from the virion surface. Each immature virion contains 60 trimers arranged forming an icosahedral symmetry (Zhang *et al.* 2003). The maturation of viral particles through the trans-Golgi network triggers the proteolytic cut of prM which results in the mature ~8 kDa M protein (Stadler *et al.* 1997). Contained inside the viral envelope layer, the ~12 kDa capsid protein (C) is associated to the viral genome forming the viral ribonucleoprotein (RNP) and anchors to the viral membrane (Ma *et al.* 2004). Viral particles do not contain NS proteins. These proteins are produced and remain in the host cell. The NS1 protein (46-55 kDa) is translocated into the ER lumen after translation, where it is fully decorated by glycosylations that allow them to bind the luminal side of ER membranes, without additional hosts proteins (Shi *et al.* 2014). NS1 is required for replication of the viral RNA and for host membrane association. The NS1 interactions with NS4A and NS4B may facilitate the recruitment of this protein (Lindenbach *et al.* 1999). NS2A (22 kDa) is a membrane protein lacking enzymatic activity. It may function as a scaffold protein, organizing the replication complex (Patkar and Kuhn 2008, Xie *et al.* 2013). NS2B is a 14 kDa membrane associated glycoprotein, involved in NS3 binding and stabilization (Luo *et al.* 2008). NS3 is a modular protein that exhibit three enzymatic activities: The N-terminal contains a trypsin-like serin protease domain, the C-terminal domain presents ATP-driven helicase activity (NTPase) for unwind the dsRNA intermediate during the genome replication and 5’ RNA triphosphatase activity (RTPase), which is a crucial step for positive sense

genome capping by NS5 (Luo *et al.* 2008, Yon *et al.* 2005). Also, it has been proposed a mechanism by which NS3 would act as an "RNA-anchoring device" avoiding re-annealing of both strands and redirecting NS5 to the 3' end of the genome (Filomatori *et al.* 2006).

Located at the most C-terminal domain of the polyprotein, NS5 is a 100 kDa protein that exhibit RNA dependent RNA polymerase activity (RdRP) and acts as the viral genome replicase in flaviviruses.

1.2.3 The NS5 protein

The NS5 protein is the largest and most conserved protein in the flavivirus genome. This protein is in charge of two essential functions. At the N-terminal end of the polypeptide, a highly conserved methyltransferase (MTase) domain is responsible for formation of the 5'-cap of the nascent viral genomic +ssRNA. This capping process is divided into four sequential steps: (a) hydrolysis of 5'triphosphate end of the nascent RNA by NS3 yielding a 5'-diphosphate end, (b) transfer of GMP from GTP to the 5'-diphosphate by the RNA guanylyltransferase (GTase) activity, also present in the NS5 N-terminal domain (Issur *et al.* 2009), (c) methylation of N-7 of guanine yielding ^{m7}GpppA-RNA, known as cap-0 and (d) further methylation of the ribose 2'-OH position of the first RNA nucleotide, finally yielding ^{m7}GpppA^{m}2'O-RNA which is known as the type I cap structure. In flavivirus, both N-7 and 2'-O methylation, are performed by the NS5 MTase domain, using S-adenosyl-L-methionine (SAM) as a methyl donor group and hence, generating two S-adenosyl-L-homocysteine (SAH) molecules as reaction by-products (Lu *et al.* 2013). Capped RNAs by viral MTases converts viral RNA indistinguishable from host mRNA, escaping in this way from the host defense machinery and increasing their stability.

The crystal structure of DENV NS5 MTase domain showed the typical S-adenosyl-L-methionine (SAM)-dependent MTase structure, in which a Rossman-fold comprised by 7-strand β-sheet flanked by a two α-helices from one side and another α-helix from the other side, harboring the SAM binding site (Egloff *et al* 2002). The active site is located on one side of the β-sheet with a K-D-K-E catalytic tetrad located in the center surrounded at the sides by the SAM and cap ^{m7}Gpp binding pockets. The MTase domain is connected to the NS5 C-terminal

domains through a long and flexible loop. Several studies have described that the MTase may adopt multiple relative positions in solution depending on the interactions with other components of the replicative complex (Bussetta *et al*. 2012, Saw *et al*., 2015).

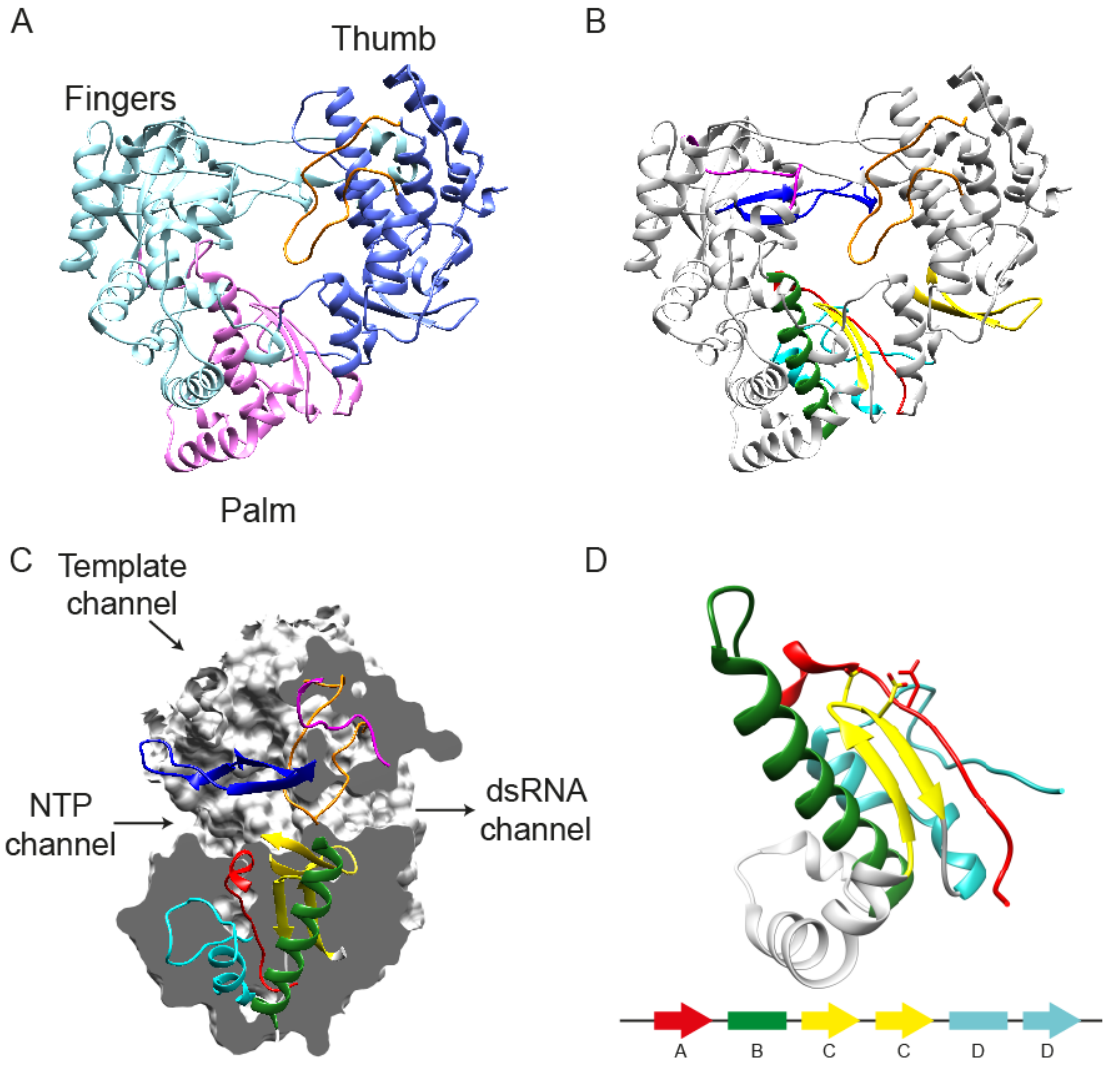


Figure 4: Overall structure of a flavivirus NS5 RdRP domain (JEV PDB: 4K6M). **(A)** The three RdRP conserved subdomain: fingers, palm and thumb are colored in cyan, pink and blue respectively. The priming loop is colored in orange **(B)** Distribution of the RdRP conserved motifs: A (red), B (green), C (yellow), D (cyan), E (wheat), F (blue) and G (purple). **(C)** Surface representation of a RdRP section with the conserved structural motifs shown as in B. The template and NTP entry channels, and the dsRNA exit channel are indicated with arrows and explicitly labeled. **(D)** Detailed view of the architecture of the RdRP palm subdomain. Cartoon of secondary structure elements are colored as in (B). Catalytic aspartic acid residues in the active site are as sticks with carbon atoms colored in yellow and red.

The NS5 C-terminus contains the RNA-dependent RNA polymerase (RdRP) domain, in charge of genome replication. In presence of metal ions, RdRPs catalyze the formation of phosphodiester bonds between ribonucleotides in a RNA-template dependent manner, whereas replication initiation in flaviriruses

occurs at the 3' end of the template strand in a primer-independent manner and continues in the 5'-3' direction.

The shape of RdRPs imaginatively resembles a closed right hand with 'fingers', 'palm' and 'thumb' domains, providing the correct geometrical arrangement of substrate molecules and metal ions at the active site for catalysis (Figure 4 A). Fingers, palm and thumb sub-domains of RdRPs collaborate to each other, supporting the binding of template RNA and NTPs. Previous structures of ternary RdRP-RNA-NTP revealed the presence of three channels in the RdRP architecture. These channels serve as the entry paths for template (template channel), for nucleoside triphosphates (NTP channel) and as the exit path for the dsRNA product (dsRNA channel) (Figure 4 B and C). (Ferrer-Orta *et al*., 2006; Ferrero *et al*., 2018). The NTP and template entry channels meet at the catalytic site located at the palm domain. The palm architecture is composed by three-stranded antiparallel β-sheet core, flanked by three α-helices, and is the most conserved feature among all known polymerases (Figure 4 D) (Gorbalenya *et al.* 2002). It contains four structural motifs, arranged in sequential order A-B-C-D-E from amino- to carboxyl- terminus that are essential for polymerase function. Motif A is located at the end of a β-strand in the central core while motif C is at the top of a β-hairpin. Three conserved aspartic acid residues, two at motif C and one at motif A, are crucial for the two metal ion catalytic mechanism (Beese and Steitz 1991) (Figure 4 B and D). A forth aspartic acid residue at motif A is involved in the discrimination of NTPs over deoxy NTPs by interacting with the ribose 2'-hydroxyl group (Appleby *et al.* 2015, Gong and Peersen, 2010, Shu and Gong 2016). Motif B forms an α-helix that packs against one strand of the β-sheet core. A conserved asparagine in (+)ssRNA viruses, participates in NTP selection forming a critical hydrogen bond with the 2'-OH of the incoming NTP. Motif D is comprised by an α-helix and a short loop that bends back around to form the fourth strand of the β-sheet core. The signature of this motif is a basic residue (lysine or histidine) and is involved in the regulation of the catalytic efficiency and fidelity. Motif E, is a β-hairpin structure that lies at the junction between the palm and thumb domains. Is usually termed "the primer grip" as it interacts with the nascent RNA at -2 and -3 positions, aiding to the correct positioning of the 3' hydroxyl group of the primer for catalysis (Appleby *et al.* 2015). The fingers sub-domain, located at the RdRP N-terminus, can be further split into: index, middle,

ring and pinky fingers and contains two conserved motifs: motif G is located in a loop outlining the template channel entry. It plays an important role in the active site configuration and holds the template RNA in a position that facilitates the polymerization. The Motif F, harbors a conserved arginine and, together with other partially conserved basic residues form the roof of the NTP channel (Figure 4B) (Ferrero *et al.,* 2018).

The architecture of the thumb domain at the C-terminus of the RdRP is poorly conserved. It harbors residues involved in template RNA packing and stabilizing the initiating NTPs of the incoming RNA template (te Velthuis 2014). This subdomain also facilitates the translocation of the template RNA through large conformational rearrangements. Flavivirus thumbs contain elements that protrude into the template channel to serve as priming platforms that stabilize the *de-novo* initiating complexes (Figure 4 A, B and C) (Appleby *et al.* 2015).

1.2.4 *De novo* initiation of RNA synthesis

Correct initiation of RNA synthesis is crucial to maintain the integrity of the viral genome. RNA viruses use mainly two different mechanisms of RNA synthesis initiation: primer dependent initiation or *de novo* initiation (van Dijk *et al.* 2004; Kao et al. 2001). RNA viruses can use either one or sometimes both mechanisms. In the *de novo* initiation, the first incoming nucleotide provides the 3'-hydroxyl, serving as a primer for the addition of the next nucleotide to form the first phosphodiester bond of the nascent strand. *Flaviviridae*, as well as other virus families, such as *Leviviridae*, *Orthomyxiviridae* and *Reovidridae* (Kidmose *et al.* 2010, Tao *et al.* 2002, te Velthuis 2016 and 2017) replicate its genome using *de novo* initiation. Structural comparisons showed that RdRPs of these viruses share a distinctive characteristic for ensuring competent replication initiation. A large loop that comes out from the thumb subdomain acts as stabilizing platform for the 3' terminus of the template RNA, and the initiating NTPs during *de novo* initiation. This loop fills most of the active site cavity (Figure 4) and also serves as a barrier preventing the chain elongation (Butcher *et al.* 2001; reviewed in Lescar *et al.* 2009 and Ferrero *et al.,* 2018). Therefore, a large conformational change, for move the initiation platform away from the active site, would be

required to allow the transition from the initiation to the elongation states of the RNA synthesis (Appleby *et al.* 2015, Butcher *et al.* 2001, Mosley *et al.* 2012).

1.2.5 Replication factories

Flavivirus, replicate in the cytoplasm of infected cells through a strong modification of the intracellular environment. A complex network of intracellular ER-derived membranous structures are created where viral replication takes place. These structures are termed vesicle-packets, and appear to be the places where several process such as RNA replication, protein storage and virus progeny budding take place (den Boon and Ahlquist 2010). In these membranous compartments, flaviviruses isolate, concentrate the viral replication factors and protect the RNA genome intermediates from host defense machineries.

Flavivirus genome replication depends on a complex orchestration of RNA-protein, RNA-RNA and protein-protein interactions. After genome cyclization, that occurs via long-range RNA-RNA interactions, both 5' and 3' genome ends lie out in the near vicinity (Figure 1). The NS5 protein first binds the 5SLA structure located at the 5'UTR end of the genome, reaching the 3' end near the NS5-5SLA complex (Filomatori *et al* 2006). As consequence of genome circularization, the 3SL structure within the 3'UTR is predicted to partially unfold, releasing a 3' ssRNA template (Figure 1) that allows the *de-novo* synthesis of the negative sense RNA chain (-ssRNA), resulting in a RNA duplex (dsRNA) intermediate. The viral NS3 protein is predicted to unwind the dsRNA intermediate by its ATP dependent helicase activity (Xu *et al.* 2019). The negative sense (-) ssRNA genome serves then as template to produce the positive sense (+) ssRNA genome (Brinton *et al.* 2014, Malet *et al.* 2008, Roby *et al.* 2014). The viral genome synthesis works in a semi-conservative way, and the newly synthesized strand displaces the old one and only the new positive-sense strand is capped in a concerted action between the NS3 and NS5 capping machinery (Dong *et al.* 2014, Saeedi *et al.* 2013, Decroly *et al.* 2012). NS5 always starts genome synthesis using *de-novo* initiation, starting with the formation of the short primer pppAG which let the conservation of both 5' and 3' flaviviral genome ends (Selisko *et al.* 2012). After the synthesis of this short primer, the initiation complex

is formed, inducing strong conformational changes in the active center of the RdRp which changes from a closed to an open state allowing the extension of the new nascent (+) strand of the viral RNA (Choi *et al.* 2009, Appleby *et al.* 2015). Newly synthetized (+) genome can be used then for packing into new viral particles, to serve for additional rounds of genome replication or as mRNA for viral protein synthesis.

1.3 Zika Virus

1.3.1 Distribution

Zika Virus (ZIKV) was first identified in a sentinel rhesus monkey in the Zika forest from Uganda in 1947 (Dick *et al.* 1952). By the mid-20th century, ZIKV virus was also identified in a handful of people in sub-Saharan Africa and Southeast Asia but these cases were rare, isolated and might have gone unnoticed. By the 21st century, this virus experimented a great expansion across the Pacific Islands and the first major outbreak took place by 2007 on Yap Island were approximately three-fourth of the population over the age of three were suspected to be infected. In this event, the main symptoms were characterized by conjunctivitis, rash and joint pain (Duffy *et al.* 2009).

In October 2013, an outbreak of ZIKV of 29.000 cases were reported in French Polynesia. In this episode, most patients exhibited mild symptoms but, one patient developed Guillain-Barre Syndrome (GBS) being the first reported case in which this virus provoked GBS 1. This syndrome is a rare autoimmune central nervous system (CNS) disease that produce neural damage, muscular weakness and even paralysis. After this event, several outbreaks took place in New Caledonia, the Cook Islands, and the Easter Island (Musso *et al.* 2014). Phylogenetic studies demonstrated that ZIKV isolates from Yap Island (2007), Cambodia (2010) and French Polynesia (2013) belongs to the same Asian lineage (Cao Lormeau *et al.* 2014).

In 2015, ZIKV was first detected in Bahia, Brazil (Campos *et al.* 2015) were patients experienced headache, joint pain, myalgia, non-purulent conjunctivitis and lower back pain. Brazilian clinicians also reported an unexpected increase in GBS cases in ZIKV infected adults. However, it is not known how ZIKV infection

triggers GBS, and how other factors also present like DENV which is transferred by the same mosquito affect in the syndrome development (Broutet *et al.* 2016). During that year, the virus expanded across the Americas reaching the United States. During that period, clinicians described clusters of microcephaly cases associated to newborn that concurred with ZIKV cases, pointing out a possible association. Due to this global threat, on February 1st 2016, the World Health Organization (WHO) declared the ZIKV epidemic a Public Health Emergency of International Concern (Heymann *et al.* 2016).

1.3.2 Transmission

Like other Arboviruses, ZIKV is transmitted by the bite of a mosquito where it is deposited in the epidermis and dermis during mosquito feeding. Initial infection likely occurs in skin immune cells including fibroblast, epidermal keratinocytes and dendritic cells (Hamel et al. 2015). Other ZIKV transmission routes have been described including perinatal transmission, sexual transfer and blood transfusion (Besnard et al. 2014, Barjas-Castro et al. 2016, Musso et al. 2015). Virus persistence in human semen can extend up to 38 days after initial detection in humans (Paz-Bailey et al. 2018) and experiments done with mice revealed that vaginal tract is highly susceptible to ZIKV replication which is highly probable also in humans (Yockey et al. 2016).

1.3.3 Pathogenesis

Normally, ZIKV infection is not fatal in adults and approximately the 80% of ZIKV cases goes unnoticed or remains asymptomatic. However, it was reported in 2016 the first case of ZIKV associated encephalitis that resulted in patient death (Soares et al. 2016) and may result fatal in immunocompromised patients (Zonneveld et al. 2016). Other complications in adults includes testis tissue damage, oligospermia, diminished testosterone and low inhibin B levels (Govero et al. 2016). In women, ZIKV also infects endometrial stromal cells (Pagani et al. 2017). ZIKV is appearing in other cells and tissues that was not seen previously in humans like urine and saliva during the acute phase of infection (Fourcade et al. 2016). It has also been reported two cases of severe thrombocytopenia (Sharp

et al. 2016). Furthermore, ZIKV infection may be associated with ocular abnormalities (Sun et al. 2016) and hearing loss (Leal et al. 2016).

However, the main concern for ZIKV lies in the ability to severely alter the neurodevelopment of newborn children of infected mothers. In 2016 it was estimated that 29% of babies born from infected mothers exhibited developmental abnormalities (Brasil et al. 2016), and tended to occur more frequently when the infection took place during the first three months (Dang et al. 2016). Studies shown that ZIKV cross the placenta and replicate in human brain cells, the risk of congenital brain abnormalities in the fetus could be around 50 times higher in mothers infected by ZIKV during the pregnancy (Broutet *et al*. 2016). In 2015, neonatal microcephaly cases increased 20-fold in Brasil, corresponding temporally and geographically to the ZIKV outbreak (Marrs et al. 2016)

The correct proliferation and differentiation of neural progenitor cells (NPCs) is essential processes for normal brain development in mammals, and disturbance of these processes result in developmental brain disorders, including microcephaly (Kriegstein et al. 2009). ZIKV has been shown to infect human NPCs (Tang et al. 2016) and induce cell death and dysregulation of cell-cycle progression in human pluripotent stem cells reducing the growth of organoid models (Garcez et al. 2016). In vivo experiments showed a link between ZIKV and microcephaly, ZIKV infection disrupted NPCs development resulting in cell-cycle arrest, cell apoptosis, differentiation inhibition, cortical thinning and ultimately, microcephaly (Li et al. 2016). Studies using ZIKV directly injected into the CNS of mice revealed virus-induced dysregulation of neural development, affecting neural differentiation in early stages of neurogenesis (van den Pol et al. 2017) which in turn leads to smaller cortex and ventricle cavities (Nguyen 2016).

1.3.3 Therapeutics

Because ZIKV and DENV share common epitopes, it was shown that anti-DENV monoclonal antibodies partially neutralized in-vitro ZIKV infection, however, subsequent in vitro and in vivo experiments in mice revealed an enhanced ZIKV infection rather than inhibition (Bardina *et al*. 2017). This data highlights the caution needed when designing flavivirus vaccines. Several DNA, inactivated or

recombinant based vaccines are in phase I studies, moreover, the National Institute of Allergy and Infectious Diseases finished recently a phase II ZIKV DNA based vaccine which consists of unmodified prM-E proteins (Ghaffar *et al.* 2018). NS5 constitutes one of the main targets for ZIKV inhibition by directly targeting virus replication, several drugs have been tested in *in vitro* and in mice. Nucleoside analogs like 7-deaza-2'-C-methyladenosine and NITD008 have been used to effectively block the NS5 activity both *in vitro* and *in vivo* experiments (Deng et al. 2016, Zmurko et al. 2016). Ribavirin, another nucleoside analog that has been historically used to treat Hepatitis C in combination with interferon, also inhibits ZIKV replication (Kamiyama *et al.* 2017). Sofosbubir, an inhibitor approved to treat Hepatitis C infections, has been shown to be an effective inhibitor of ZIKV by directly blocking the NS5 RdRP activity acting as a chain terminator (Fung *et al.* 2014). The treatment results in a reduced viremia and in a decrease in vertical transmission. Other compounds such as the broad-spectrum nucleoside analogue BCX4430, avoids ZIKV infection by inhibiting viral replication also acting as a non-obligate RNA chain terminator (Julander et al. 2017, Warren et al. 2014).

In addition to antivirals targeting the NS5 RdRP activity, the inhibitors of the protease activity of NS3 also constitutes an attractive route against ZIKV. Several drugs have been found, including the compound NSC157058, a potent inhibitor that binds the NS2-NS3 protease interface, altering the interactions between the protease and its cofactor (Shiryaev *et al.* 2017). Other compounds, like Temoporfin, nitazoxanide and niclosamide act efficiently inhibiting NS3 protease activity (Sacramento *et al.* 2017). Finally, Viperin is an antiviral protein that inhibits ZIKV replication by targeting NS3 for proteasomal degradation (Panayiotou *et al.* 2018).

1.3.4 Regulation of NS5 activity

In addition to its own RdRP and MTase activities, the interactions between these two domains also serve to regulate the different NS5 functions. In addition, NS5 shows a number of essential interactions with other viral and host proteins. The regulatory effect of the MTase domain in the NS5 RdRP activity has been a longstanding question. *In vitro* studies indicated stimulatory effects by the MTase

domain on the de-novo RNA synthesis and specially for the chain elongation process, indicating a regulatory role of the MTase domain in the overall RdRP activity of NS5 and hence, in viral replication (Zhao et al. 2017). Other studies also pointed up key interactions between MTase domain and RdRP. MTase residues E112, P113 and L115 interact with F466 residue on the RdRP domain stabilizing the F motif. Point mutations on these residues resulted in lowered RNA elongation activity but did not show effect on replication initiation (Rusanov et al. 2018).

In addition to playing a major role in the genome replication, NS5 also regulates other viral protein activities. In vitro experiments with NS3 showed a low efficiency unwinding dsRNA, however, this activity was enhanced in presence of NS5. The disruption of the NS5-NS3 interaction surface resulted in a defected viral replication, revealing the stimulatory role of NS5 on unwinding NS3 activity in viral replication (Xu et al. 2019).

NS5 also plays an important role during the viral infection by directly affecting host immune defense mechanisms. Several studies have described the disrupting role of NS5 in the interferon (IFN) signaling pathway in the host cell. During viral infection, cytokine secretion represents an effective response against the viral fight that are secreted after the activation of the IFN signal pathways. ZIKV NS5 protein differentially regulates type 1, 2 and 3 IFN signaling pathways (IFN1-3). IFN1 and IFN3 present strong antiviral effects and these are suppressed by NS5 protein, while IFN2 is activated during ZIKV infection (Chaudhary et al. 2017).
IFN1 and IFN2 pathways are involved in the triggering of the adaptive immune system for producing antibodies against ZIKV. NS5 interferes these signal pathways by direct binding of TBK1, an intermediate protein kinase of this cascade (Lin et al. 2019). IFN1 and IFN2 are also regulated by the transcription activator STAT2, promoting the INF mediated antiviral response. ZIKV NS5 protein is also able to bind STAT2 which results in proteasomal degradation in humans and hence promoting viral infection (Grant et al. 2016, Kumar et al. 2016). IFN2 signaling activates the production of pro-inflammatory cytokines and also the up-regulation of the ZIKV entry factors AXL, Tyro3 and DC-SIGN. The

activation of this pathway induce cellular inflammation, autophagy and cell death and therefore, spreading viral particles to infect the surrounding cells (Chaudhary et al. 2017)

1.4 Cryo-Electron Microscopy

Structural biology lends eyes to the life sciences, and has built the groundwork to our understanding of key processes in biomedicine and biotechnology. So far, one of the most exciting development of the last five years has been the irruption of single particle cryo-electron microscopy (Cryo-EM) into the atomic resolution realm, previously owned almost exclusively by macromolecular crystallography (Cheng, Glaeser and Nogales 2017). This “revolution in resolution” of the cryo-EM has been possible due to recent advances in instrumentation and software. Ten years ago, several virus structures were solved by cryo-EM, achieving quasi-atomic resolution (Jiang et al. 2008, Yu et al. 2008, Zang et al. 2008) and also the large ribosome structure was solved using this technique (Villa et al. 2009, Seidelt et al. 2009). All these assemblies showed a large size since these particles producing a good signal-noise projections. Also the presence of icosahedral symmetry in the case of viruses, helps in the reconstruction process. However, despite the electron microscopes would theoretically provide enough information to achieve high resolution of protein structures smaller than 100 kDa, the reality was that the structures did not reached that resolution (Henderson 1995). This problem may be attributed to the low efficiency in electron detection. Traditionally, data acquisition was performed in photographic films. Later on, developments on charged-coupled devices detectors (CCD detectors) allowed the data acquisition in an automatic way. However, detection efficiency of both methods was very poor. In 2012, several companies produced the firsts direct electron detectors (DED) that allowed the detection of one single electron event which dramatically improved the detection efficiency compared to the previous methods and also the technology behind DED allowed faster data acquisition (McMullan et al. 2014). When the electron beam hits the vitrified sample, the thin layer of water embedded sample experiments the so called “beam induced movement” associated to the high energy beam. This movement produces a blurring effect on the micrographs that difficult the subsequent reconstruction.

With the new DED technology, this problem is overcome by dividing the total exposure time into tens of recorded frames in a “movie-like way”. Thus the movement of each particle is recorded several times along the whole displacement allowing the correction with posterior processing (Campbell et al. 2012), this resulted in a dramatic increasing on the overall structure resolution.

Biological macromolecules may exhibit a highly dynamic behavior or be composed by different subunits (which in turn may also move). This results in heterogeneous samples difficult to purify because the presence of different species. Single particle reconstruction (SPR) cryo-EM takes advantage of the presence of several species in solution. The newly developed maximum-likelihood methods allow the separation of the 2D projections in a robust and reliable way as it incorporates statistical parameters that validate the generated 3D reconstructions (Fernandez-Leiro and Scheres 2006). This methodology allows to “catch” different macromolecular states in one single experiment that would be impossible by using other structural methods. In 2013, the DED were launched into market entailing the “resolution revolution”. The resolution increasing leaded to better assignation of particle orientations, as well as the capacity to classify particles experienced a big step forward with the implementation of the Bayesian approach to the maximum-likelihood method (Scheres et al. 2012). As result, a large number of laboratories started working in cryo-EM SPR, solving a large number of high resolution structures (Bai et al. 2015; Tortorici *et al.* 2019).

The new developed tools for solving macromolecular structures have great capacity of separation of homogeneous particle populations, allowing the reconstruction of 3D maps with just few thousands of 2D projections. This is particularly interesting when looking structures of complexes in which just a small population of the molecules are in an interesting special conformation or when labile complexes are studied.

Recently, in a large number of articles reported in high profile journals, the two fundamental barriers dampening cryo-EM have been overcome: crossing beyond the 2 Å resolution barrier (Merk et al. 2016) as well as obtaining structures of

proteins with sizes under 100 kDa (Herzik et al. 2019), establishing that cryo-EM can be used generally in the detailed investigation of a broad spectrum of samples, ranging from large multicomponent complexes down to drug-target interactions (Renaud et al. 2018) and dynamic conformational states (Nguyen et al, 2016).

2. Objectives:

1 Structural and biochemical characterization of the ZIKV NS5 protein

-Cloning, expression and purification of NS5.

-Crystallization and structure determination of NS5 using X-Ray crystallography.

-Characterization of the NS5 dimerization and multimerization capacities, using combined biochemical and biophysical techniques.

-Evaluation of the NS5 effect on the normal neural tube development in chicken embryos.

2 – Structural and biochemical characterization of NS5 in complex with the 5SLA promoter.

-Production of the 5SLA RNA by *in-vitro* transcription and purification for structural and biochemical studies.

-Evaluation of the NS5-5SLA binding

-Optimization of the NS5-5SLA complex formation for cryo-EM studies, using biochemical and biophysical techniques.

-Structure determination of the NS5-5SLA complex, using cryo-EM and 3D image reconstruction

-Identification and evaluation of the NS5-5SLA interfaces

-Evaluation of the NS5 *de-novo* polymerase activity when bound to 5SLA.

3. Results

3.1 Structural and Biophysical characterization of ZIKV NS5

3.1.1 NS5 expression and purification

The recombinant ZIKV NS5 protein (Suriname strain isolate Z1106033; residues 1–904) was produced in *Escherichia coli*. The highest levels of soluble protein were observed when cell cultures were incubated at 18 ºC and induced with 0.1 mM IPTG (Figure 5A). The ~100 KDa protein was purified to homogeneity after stablished an optimized protocol that turned out in the combination of a Ni^{2+} affinity column and two additional chromatographic steps: cation exchange (HiTrap SP) and size exclusion (Superdex 200 10/300) columns (Figure 5B-G).

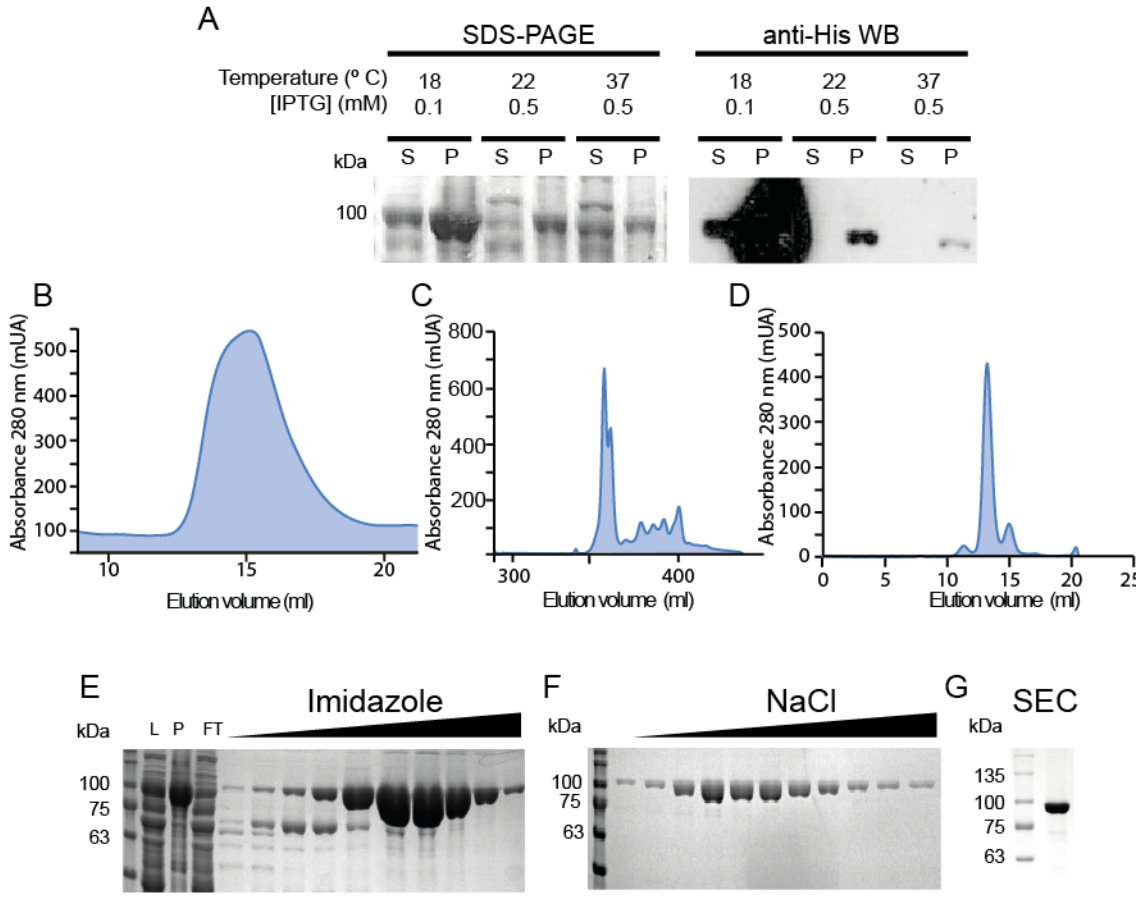


Figure 5: ZIKV NS5 expression and purification (A) Expression and solubility assays of ZIKV NS5 using different IPTG concentrations and temperatures. Lanes labeled as S correspond to the soluble fraction while P correspond to the insoluble fraction. Best conditions were obtained inducing 0.1 mM IPTG and incubating at 18 ºC. **(B)** – **(D)** Chromatograms obtained for the IMAC Ni^{2+} Affinity, Cation exchange and Size exclusion columns respectively. **(E) - (G)** SDS-PAGE analysis of the eluted samples from (B) – (D) purifications respectively. Labels above the lanes correspond to L: total lysate, P: pellet and FT: column flow through.

3.1.2 Crystal structure of NS5

The NS5 protein was crystallized in two different crystal forms: orthorhombic $P2_12_12_1$ and hexagonal $P6_5$ (Figure 6). The first diffraction data collected was at very low resolution (> 6 Å) and several optimization trials were performed including crosslinking, capillary crystallization or seeding. However, all these optimization attempts resulted unsuccessful. Finally, after applying a soft dehydration protocol by soaking in harvesting buffer that included a 5% more of the phosphate salts in the reservoir solution, we obtained crystals that diffracted at 4 Å and 4.8 Å resolution for the $P2_12_12_1$ and $P6_5$ space groups respectively, and allowed us to solve the structure. Orthorhombic $P2_12_12_1$, with unit cell parameters of a = b = 234.58 Å, c = 406.12 Å and hexagonal $P6_5$, with unit cell parameters a = 191.06 Å, b = 192.06 Å, c = 407.23, contained six independent NS5 molecules in each asymmetric unit (Matthews coefficient (Vm) of 5.38 and 6.32 corresponding to a solvent content of 77% and 80% for the orthorhombic and hexagonal space groups, respectively) (Table 1) (Figure 7).

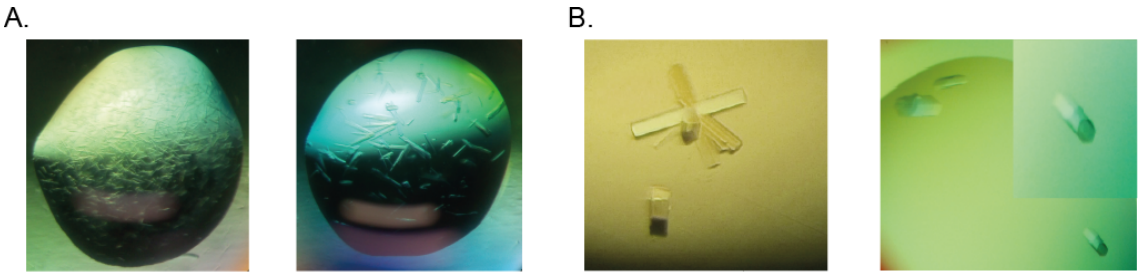


Figure 6: NS5 crystallization **(A)** First NS5 crystals obtained by vapor diffusion in hanging drops. Crystals grew in 0.1 M Sodium Acetate, 0.8 M Sodium di-hydrogen phosphate, 1.2 Di-Potassium hydrogen phosphate pH 4.5 before (left drop) and after optimization (right drop). **(B)** Different crystal forms of NS5 protein obtained in sitting drops by mixing 300 nl of mother liquor with 300 nl of protein solution at 10 mg/ml. Orthorhombic crystals (space group $P2_12_12_1$) are shown in the left panel. The right panel shows hexagonal crystals (space group $P6_5$).

Both crystal forms showed important anisotropy. In fact the analysis of the $P2_12_12_1$ data, using the STARANISO software (Tickle *et al.* 2018), revealed diffraction limits of ellipsoid fitted to diffraction cut-off surface of: 5.85, 7.44, and 3.76 Å along the a*, b*, and c* axes. The resolution cut-off was calculated using the following parameters: Rpim = 0.6, I/σI = 2.0 and CC1/2 = 0.3. The application of the anisotropy correction to the unmerged data produced a best resolution limit

of 3.98 Å in direction 0.071 a* +0.010 b* + 0.997 c* and a lowest cut-off limit of 7.39 Å in direction 0.857 b*+0.514c* (Table 1).

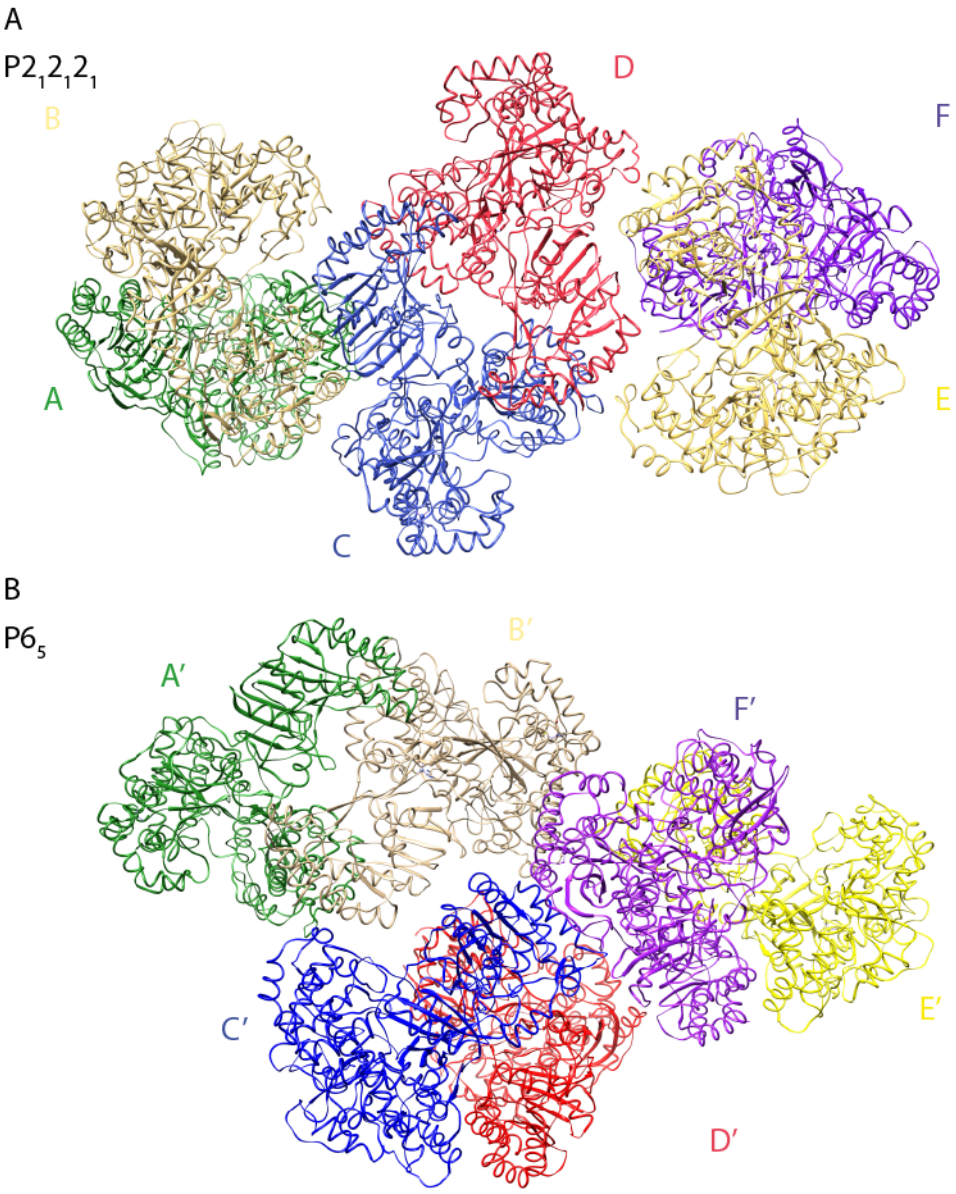


Figure 7: Ribbon representation of the crystallographic asymmetric unit corresponding to the **(A)** $P2_12_12_1$ and **(B)** $P6_5$ crystal. The six monomers are labeled from A-F (A in green, B in brown, C in blue, D in red, E in yellow and F in purple) and are arranged forming three homodimers in both crystal forms.

Structures were solved by molecular replacement method, using two independent search models derived from the ZIKV MTase (PDB: 5KQR) and JEV RdRP (PDB: 4K6M) domains. In both crystal forms, the asymmetric unit is composed by six independent copies arranged in three quasi-equivalent dimers (Figure 7). Despite the limited resolution of both datasets, the molecular

replacement models showed a perfect fitting into the electron density maps and even allowed to trace bulky side chains.

Table 1. Data collection and Refinement Statistics

Parameters	5M2X	6I7P	5M2Z
Data Collection			
Resolution range (Å)	49.52 – 4.99 (5.11-4.99)[#]	49.52-3.98 (4.05-3.98)[#]	44.33-4.8 (5.06-4.80)[#]
Lowest cut-off difraction limit (Å) (direction)[¶]		7.39 (0.857b* + 0.514c*)	
Best diffraction limit after cut-off (Å) (direction)[¶]		3.98 (0.07a* + 0.01b* + 0.99c*)	
Space group	$P2_12_12_1$	$P2_12_12_1$	$P6_5$
Unit cell parameters a, b, c (Å) α, β, γ (º)	191.06 192.06 407.23 90 90 90	191.06 192.06 407.23 90 90 90	234.58 234.58 406.12 90 90 120
Measured reflections	1,962,662 (224755)	3,890,662 (143079)	202753(27952)
Unique reflections	33882(4342)	128656 (6433)	61422 (8906)
Rmerge[§]	0.750(1.694)	0.963(1.941)	0.159 (1.495)
Rpim[◊]	0.097(0.273)	0.175 (0.402)	0.104(1.005)
Multiplicity	57.9(51.8)	30.2(22.2)	3.3 (3.1)
Completeness (%)	99.5(97.3)	99.6 (92.0)	99.7 (99.4)
Mean I/Sigma	5.8(3.6)	5.2(2.0)	5.9 (1.1)
Refinement			
Resolution range	49.52 – 4.99	49.52-3.98	44.07-4.80
Rwork[†]	0.2270	0.2805	0.2410
Rfree[‡]	0.2760	0.2805	0.2660
Protein residues	5297	5297	5298
Ligands	14	14	14
Rms bond lenghts (Å)	0.0112	0.004	0.0104
Rms angles (º)	1.3744	0.665	1.2571
Ramachandran plot (%) Favored	94.66	94.53	95.67
Allowed	4.88	3.88	3.76
Outliers	0.45	0.11	0.57

[#] The number in parentheses refers to the last (highest) resolution shell.
[¶] These statistics are for data that were truncated by STARANISO.
§ Rmerge = Σ|I j - <I>| / Σ I j where I j is the intensity of an individual reflection and <I> is the average intensity of that reflection.
† Rwork = Σhkl ||Fobs(hkl)|—|Fcalc(hkl)|| / Σhkl |Fobs(hkl)|, where Fobs and Fcalc are the structure factors, deduced from measured intensities and
calculated from the model, respectively.
‡ Rfree = as for Rwork but for 5% of the total reflections chosen at random and omitted from refinement
◊ Rpim = $\sum_{hkl} \sqrt{1/n-1} \sum |I_i(hkl) - \bar{I}(hkl)| / \sum \sum I_i(hkl)$

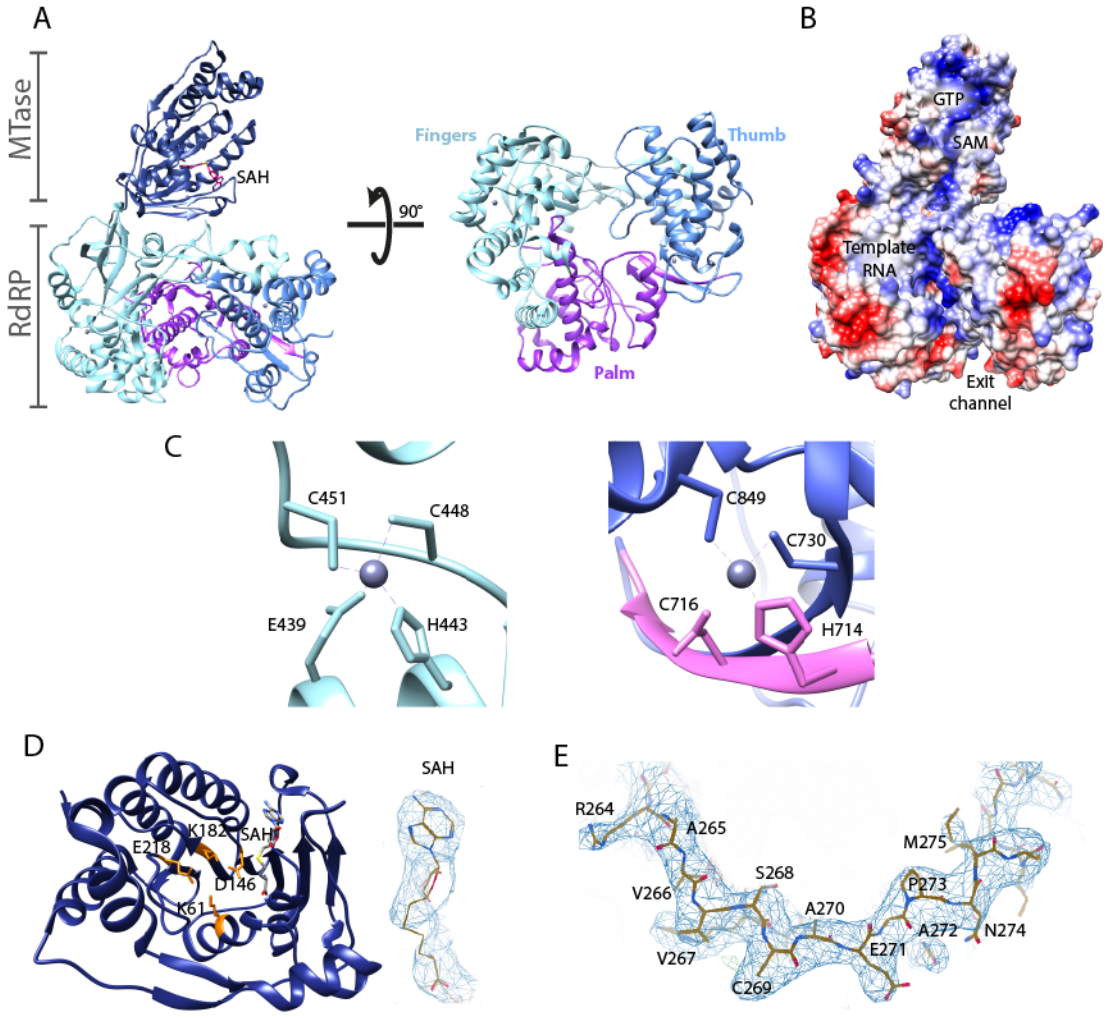


Figure 8: The structure of the full-lenght ZIKV NS5 protein. **(A)** In ribbon diagram of a NS5 monomer, the MTase domain is colored in dark blue with the bound SAH cofactor in red (left panel). The center panel shows a front view of the RdRP domain. The different RdRP subdomains fingers, pal and thumb are colored in cyan, pink and blue respectively. **(B)** Electrostatic potential surface representation of NS5 colored in blue and red for positively and negatively charged atoms respectively. The GTP and SAM/SAH binding pockets at the MTase domain and, the RNA template entry channel and the dsRNA exit channel are explicitly labeled. **(C)** Detailed view of the two Zn^{2+} metal atoms present at the RdRP domain. Coordinating residues are explicitly labeled. **(D)** A detailed view of the MTase domain where the Rossman fold can be clearly seen. Atoms that forms the catalytic tetrad are colored in orange and respectively labeled. At right, a view of the X-Ray electron density map that allowed the positioning of SAH cofactor. **(E)** Detailed view of the X-Ray electron density map (weighted 2Fo-Fc, 1.1 σ level) corresponding to the linker from residues 264 – 275.

The overall fold that ZIKV NS5 adopts is likely similar to those observed in other flavivirus NS5 crystal structures (Figure 8A). Located at N-terminal, the MTase domain is complexed with S-adenosyl-L-homocysteine (SAH) and folds in the back part of the RdRP domain (Figure 8 A and B). A large loop (residues 263-272) following the MTase connects the RdRP domain, it is likely mobile allowing the relative movement between both domains as described by Busseta *et al*. 2012. However, it could be traced in the electron density maps in four out of the

six molecules in the asymmetric unit (Figure 8E). Both MTase and RdRP are in closing contact through a 950 Å^2 interface. The NS5 most C-terminal region (residues 889-913, that contained the last 15 residues plus the 6 His-tag and the spacer) is probably flexible as it could not be traced in the electron density maps.

Located at the C-terminal moiety, NS5 RdRP domain comprises the residues spanning from 273 to 888. It adopts the characteristic closed right hand conformation as others RdRP (Applebay *et al.* 2015, Venkataraman *et al.* 2018) proteins that consists in three subdomains: thumb, palm and fingers (Figure 8A and Figure 9).

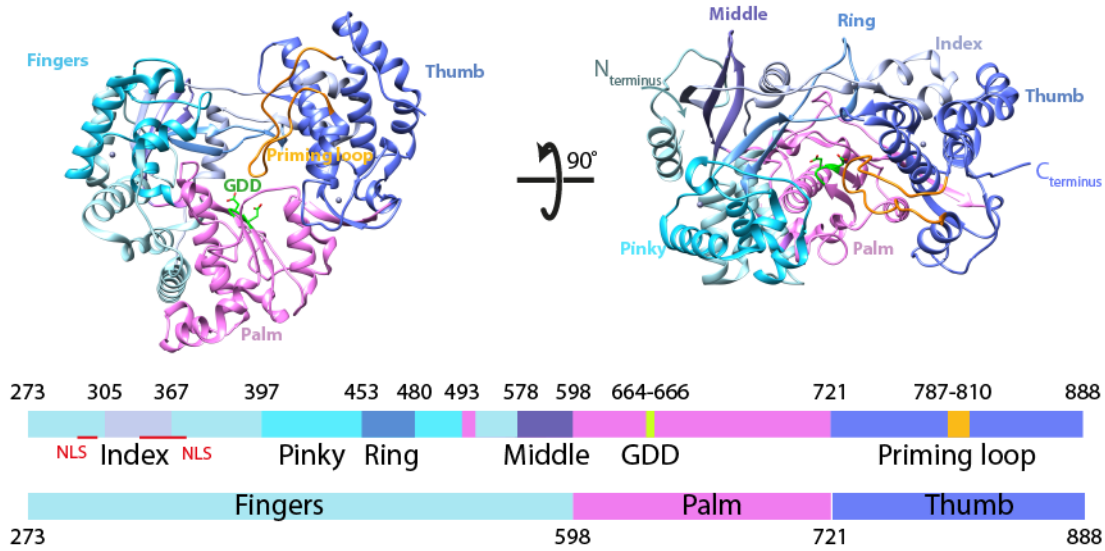


Figure 9: Cartoon representation of the different structural features of NS5 RdRP domain. At left, GDD motif in palm subdomain is colored in lemon green. Priming loop in thumb subdomain is colored in orange. At right, RdRP top view where the index, middle, ring and pinky fingers are colored in different blue tones. Bottom bar indicates the residue numbers of depicted features as well as the three RdRP subdomains. Nuclear localization signals are indicated as “NSL” in red.

In ZIKV NS5, the thumb domain is bulkier than others RdRPs as it allows *de novo* initiation synthesis with an accessory protruding loop that inserts directly into the active site cavity and serves as a priming platform for the polymerization of the first nucleotide using a conserved aromatic residue (Selisko *et al.* 2014). This structure is termed “priming loop” and comprises the residues 787 to 810 placed between two α-helix (Figure 9). In ZIKV NS5, the tip of this loop contains two residues, Trp797 and R794, contacting through cation-π interactions in a suitable orientation for priming the initial NTP as previously described in the JEV NS5 structure (Lu *et al.* 2013). This structure is likely flexible and might experiment

conformational changes as the newly synthetized RNA strand moves forward through the central channel. The palm subdomain is placed in the bottom-middle part and connects both fingers and thumb subdomains. NS5 motif C appears conserved and harbors the catalytic GDD motif (residues from 664-666) flanked by two antiparallel β-strands (Figure 9). NS5 fingers subdomain is divided into index (residues 305 - 367) that connect with the thumb subdomain, the middle (residues 578 - 598), the ring (residues 453 - 480) is forming the roof of the NTP entry channel and the pinky (residues 397 – 493) (Figure 9).

MTase domain (residues 1-262) adopts a classical Rossman fold composed by a central seven-stranded β-sheet surrounded by several α-helices (Figure 8D). This folding is consistent with others observed in flavivirus like JEV, DENV or WNV (Lu *et al.* 2013, El Sahili *et al.* 2019, Zhou *et al.* 2007) and sequence is largely conserved in other flaviviruses (Figure 10). MTase active site is composed by the catalytic tetrad K61-D146-K182-E218 which is conserved to other described flavivirus MTase (Lim *et al.* 2016, Lu *et al.* 2013). SAH cofactor is a byproduct of MTase reaction. Copurified SAH from the bacterial host could be traced into the cofactor binding pocket in two of six molecules in the $P6_5$ crystals while it was visible in all copies in the $P2_12_12_1$ crystals (Figure 8D). Co-purified SAH was also reported in other flavivirus NS5 crystal structures of ZIKV, JEV and DENV (Upadhyay *et al.* 2017, Lu *et al.* 2013, Klema *et al*., 2016, Zhao *et al.* 2015) and MTase crystal structures of ZIKV, DENV, JEV (Coloma *et al*., 2016, Egloff *et al*., 2002;) occupying the same position. It was also described in other viral methyltransferases like in Bluetongue virus or Vaccinia virus (Sutton *et al.* 2007, De la Peña *et al.* 2007). The two Zn^{2+} atoms previously found in other flavivirus RdRP are also present in ZIKV RdRP (Figure 8C); one atom is bound to the pinky finger and is coordinated by the residues E439-H443-C448-C451. The second is located between the thumb and motif E in the palm subdomain and is coordinated by H714-N716-C730-C849 (Figure 8C). In addition, three strong electron density peaks seen in the MTase NTP binding site and in the RdRP template and nucleotide binding channels respectively were interpreted by the presence of phosphate ions present in the crystallization condition. A nuclear localization signal (NLS) can be found in the finger subdomain (Figure 9) and partially overlapping the index finger (Zhao *et al.* 2017).

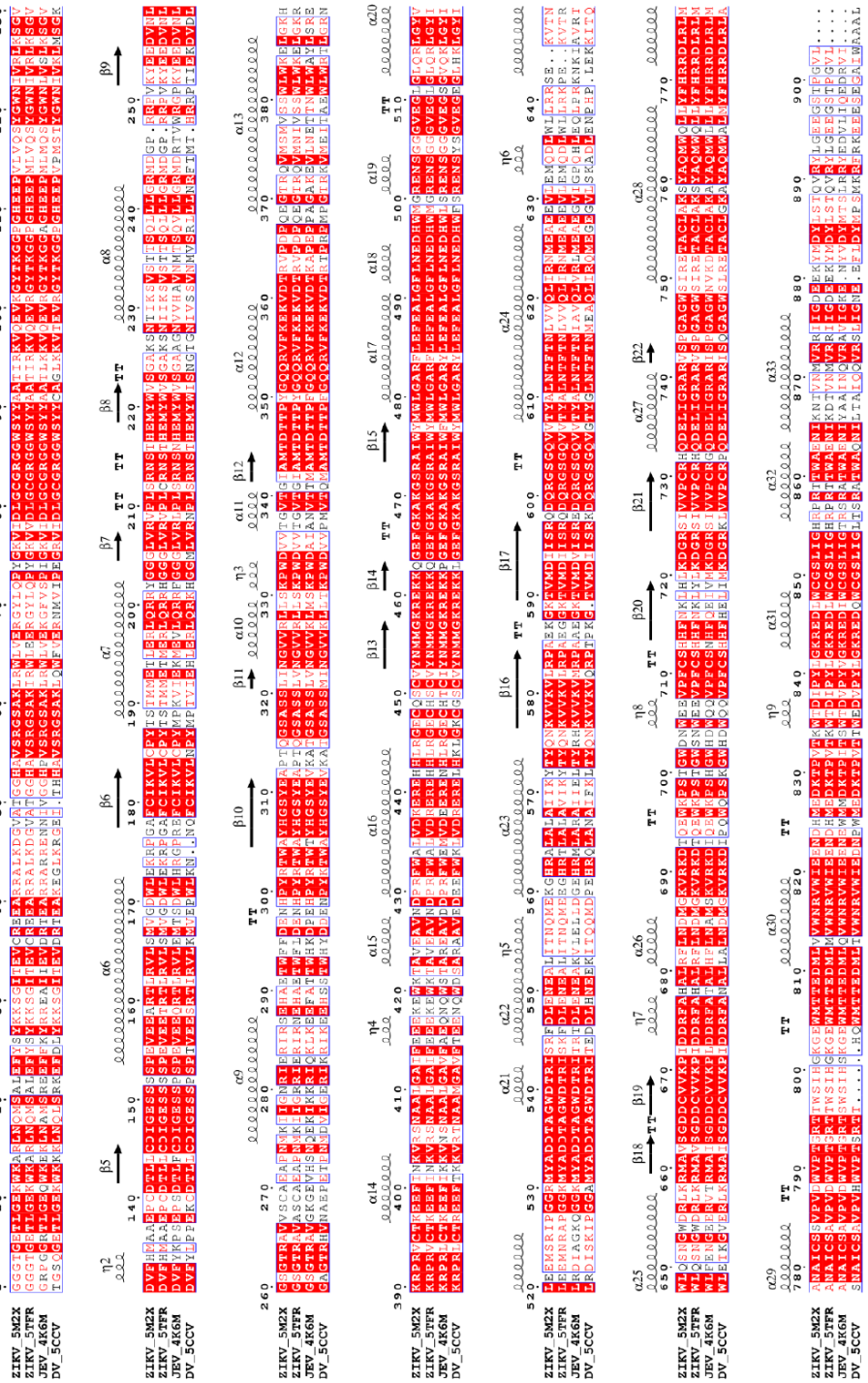


Figure 10: Sequence alignments of different flaviviral NS5 proteins: Suriname ZIKV human isolate (PDB: 5M2X, 5M2Z described in this work), ZIKV strain MR776 (PDB: 5FTR) JEV (PDB: 4K6M) and DENV (PDB: 5CCV). White letters in red boxes indicate identity, and red letters residue conservation. Secondary structure definitios, as determined in the crystal structures of ZIKV NS5 solved in this work are shown above the sequence. α-helices (α) or 3_{10} helix (η) are indicated with helical lines, β-strands are indicated with black arrows and β-turns are indicated with TT.

The relative disposition of the MTase and RdRP domains (Figure 11) is almost identical in all twelve independent NS5 molecules included in the asymmetric units of both $P2_12_12_1$ and $P6_5$ crystal forms, showing only small hinge rotations (superimposition of the twelve monomers yields rmsd values ranging from 0.1 to 0.6 Å and hinge rotations of 5° to 7°). Comparisons with the NS5 structures currently available for JEV (PDB: 4K6M), DENV2 (5ZQK, 6KR2 and 6KR3) and DENV3 (5CCV) (Lu *et al* 2013; el Sahili *et al.* 2019; Wu *et al.* 2020) show that the ZIKV and JEV NS5 structures are easily superimposable (rmsds from 1.0 to 1.3 Å), as well as the DENV (6KR2 and 5ZQK, rmsds from 0.64 to 3.05). However, despite the high degree of sequence conservation, both the relative orientation and intermolecular interactions between the MTase and RdRP domains in these viruses differ significantly from other published structures of DENV (5CCV and 6KR3) (Figure 10 and 11).

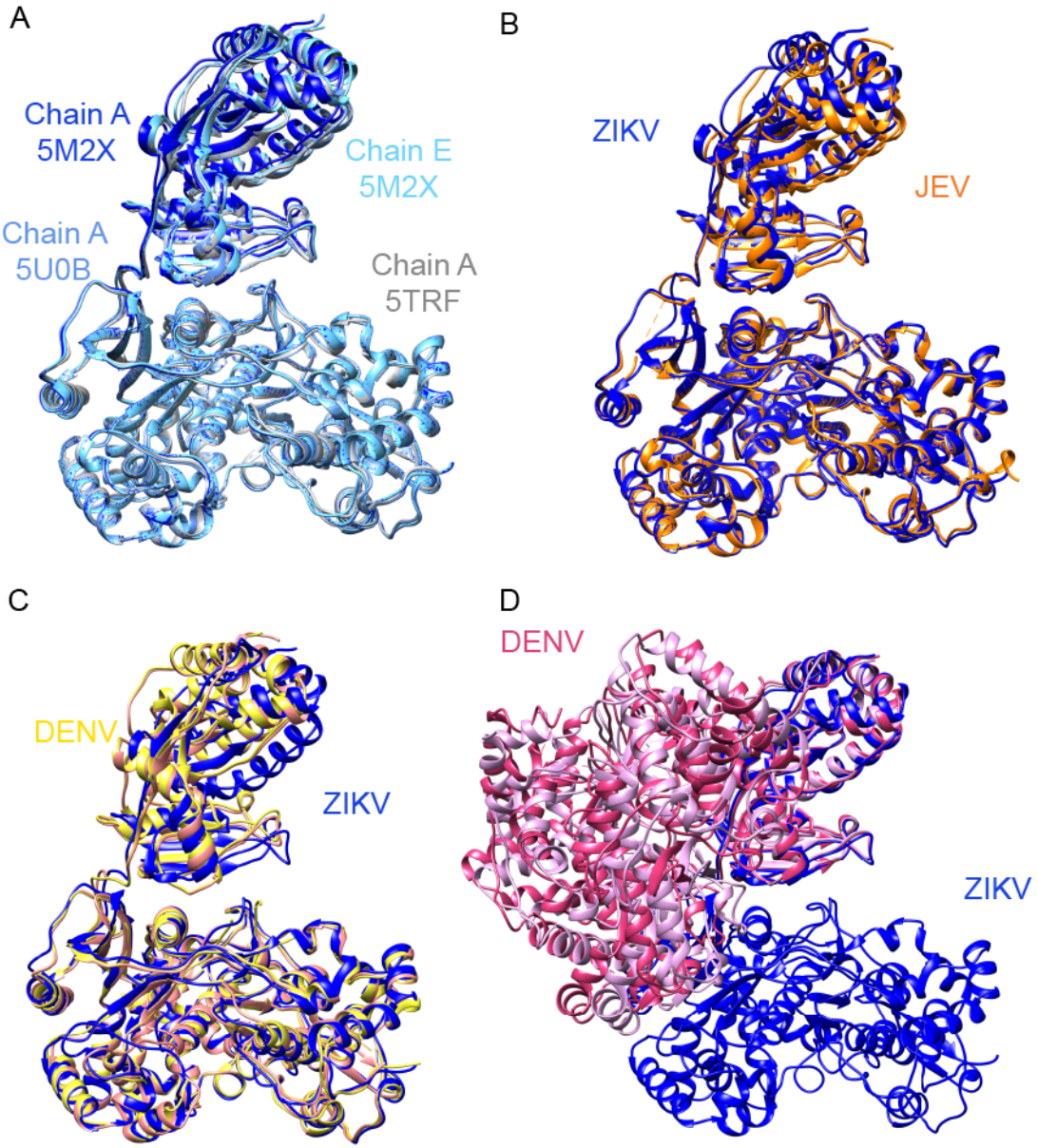


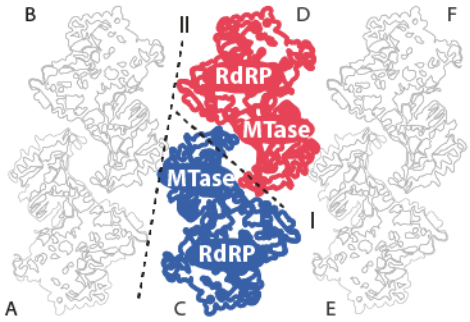
Figure 11: Structural comparison of different flavivirus NS5 proteins. **(A)** Structural alignment between the RdRP domains of ZIKV NS5. The coordinates of four different molecules are superimposed, chains A (blue) and E (cyan) in the $P2_12_12_1$ crystal structure (PDB: 5M2X; this thesis), chain A (grey) of the $P2_12_12_1$ structure (PDB: 5FTR) and chain A (pale blue) of the $P2_12_12$ crystal structure (PDB: 5U0B), showing the relative movements of the MTase domains. **(B)** Structural alignment between the RdRP domains of ZIKV, chain A (blue) with JEV NS5 (PDB: 4K6M; chain A, orange). **(C)** Superposition of the RdRP domains of ZIKV, chain A (blue) with two DENV NS5 structures (PDB: 6KR2; chain A, yellow and PDB: 5ZQK; chain A, light orange). **(C)** Structural alignment between the MTase domain of ZIKV NS5 (blue) and two DENV NS5 structures (PDB: 5CCV, chain A, magenta and PDB: 6KR3, chain A, pink).

3.1.3 Quaternary arrangement of ZIKV NS5

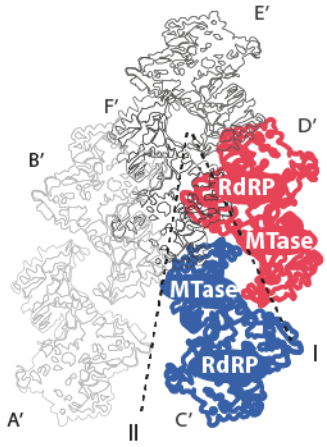
The crystallographic asymmetric unit of both crystals consists of three dimers (named AB, CD, and EF in the crystal $P2_12_12_1$ and A'B', C'D', E'F' in the crystal

$P6_5$) (Figure 7) in which the NS5 monomers are related a by pseudo-twofold molecular axes. Each NS5 monomer contributes >1200 $Å^2$ (30% of the total surface) to the dimer interface (Figure 12).

A



B



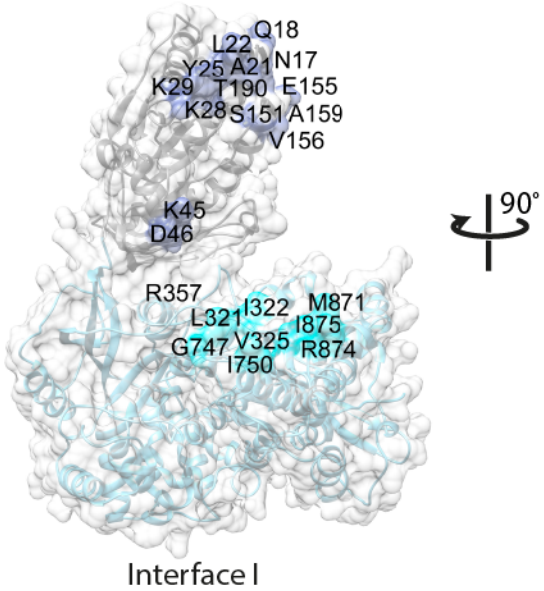
C

Interface	Chain	Area $(Å)^2$
I	AB	1406.1
	CD	1277.6
	EF	1220.1
II	AB + CD	1086.1
	CD + EF	1106.7
	EF + AB	1035.9

D

Interface	Chain	Area $(Å)^2$
I	A'B'	1513.0
	C'D'	1586.5
	E'F'	1609.5
II	A'B' + C'D'	911.5
	C'D' + E'F'	886.6
	E'F' + A'B'	235.5

E



F

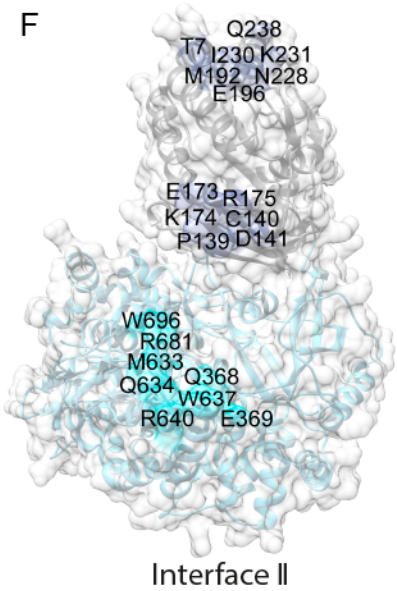


Figure 12: Schematic representation of the arrangement of the three dimers present in the **(A)** $P2_12_12_1$ crystals and in the **(B)** $P6_5$ crystals. The reference monomers are colored in red and blue and monomer-monomer and dimer-dimer interfaces are indicated with dashed lines. **(C)** Table sumarizing the buried surfaces for the $P2_12_12_1$ crystals and **(D)** for the $P6_5$ crystals. **(E)** Surface representation of a NS5 monomer. Residues involved in monomer-monomer contact (Interface I) are explicitly labeled. Residues located within MTase are colored in dark blue while RdRP residues are colored in cyan. **(C)** Surface representation of a NS5 monomer. Residues involved in monomer-monomer contact (Interface II). As in **B**, labels and color codes indicate the interacting residues.

The interface of interaction (interface I) involves two types of intermolecular contacts MTase-MTase and MTase-RdRP. MTase-MTase interactions involve the periphery of the MTase of one NS5 molecule: helices α1(N17,Q18), α2(A21, L22, Y25, K28) and the α2-β1 loop (K29) which contact residues K45 and D46 (α-3 helix) of the second NS5 subunit. MTase-RdRP contacts include the MTase α6 residues, E155, V156, A159 of the first monomer which contact two RdRP subdomains of the neighbor NS5 involving residues: a) L321, I322, V325 (β11) and G324, V325 (α10) within fingers and b) G747, I750 (β22-α28 loop) and M871, R874 and I875 (α33) at the back of the thumb (Figure 12E).

In this arrangement the MTase active site cavities of the two interacting molecules are directly connected and the two RdRP exit channels remain accessible to the solvent. The six dimers can be superimposed with rmsd values ranging from 0.8 Å to 1.0 Å for the superimpositions of all 1,776 equivalent Cα atoms. Similar head-to-tail dimers were also observed in the structure of JEV NS5 (PDB: 4K6M) but showing a slightly different rotation between monomers.

Dimer-dimer interactions (Interface II) are also conserved in both ZIKV NS5 crystal forms and involve contacts between molecules B-C and D-E ($P2_12_12_1$ crystals) and B'-C' and D'A'' ($P6_5$ crystals, with A'' corresponding to a symmetry related A' molecule). Three distinct NS5 regions are implicated in interface II contacts, burying >1000 $Å^2$ of surface area between monomers in each pair (Figure 12F). Similar contacts were observed between the two molecules contained in the asymmetric unit of the African ZIKV strain: MR776, NS5 crystals (PDB: 5TFR; 5TMH; 5U0B).

Due to the NS5 dimer distribution within the both crystal packing, it can be observed a helical distribution that forms a fibril-like structure along the c-axes (Figure 13).

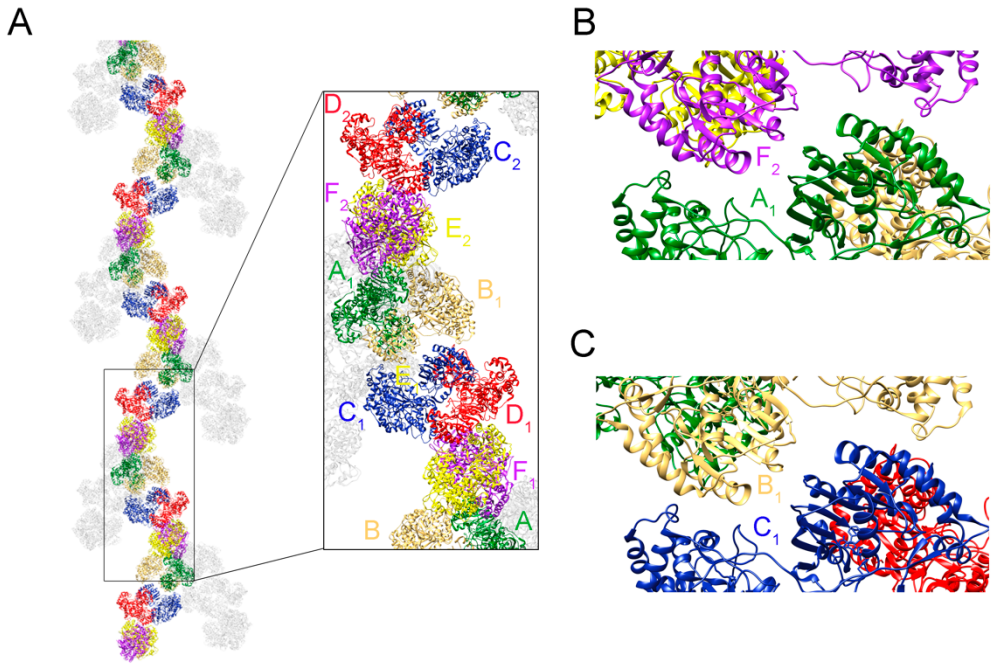


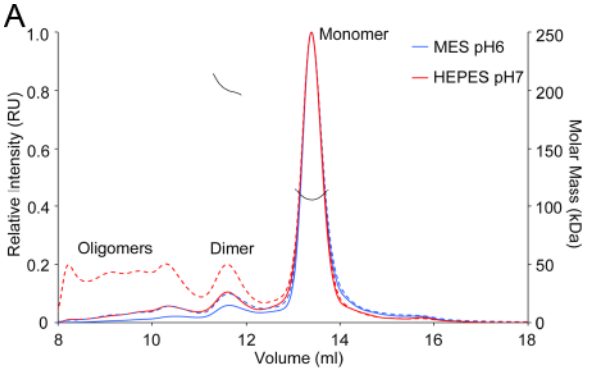
Figure 13: Contact interactions along the crystallographic c axis in the $P2_12_12_1$ crystals. **(A)** Organization of the 3 dimers of the asymmetric unit to form long fibers along the unit cell axis c. Panels **B** and **C** panels show the contact interfaces.

3.1.4 NS5 dimers are present in solution

The conserved dimer based, helical structures seen in both crystal forms, suggested that these oligomers could be stable arrangements that might exist in solution and possibly in cells. Size exclusion chromatography (SEC) in a temperature controlled chamber (25 ºC) and coupled to a multi angle light scattering system (MALS), showed that the purified NS5 construct behaved mainly as a monomer in the protein stabilization buffer, containing high salt content (50 mM MES pH 6.0, 500 mM NaCl, 10% glycerol, 5 mM DTT). A predominant peak (95 ± 2% sample mass), corresponding to 104 ± 2 kDa matches the calculated molecular weight of this protein. However, a second peak (4 ± 2% sample mass) corresponding to 178 kDa +/- 2 kDa was also present, revealing the presence of NS5 dimers in solution. Similar results were obtained with an alternative NS5 stabilization buffer at different pH but same ionic strength (50 mM HEPES pH 7.0, 500 mM NaCl, 10% Glycerol, 5 mM DTT). The

predominant peak (89.5 ± 0.1%), corresponding to a molecular weight of 105.8 ± 0.3 kDa eluted after less abundant species that included a second peak of 202.3 ± 0.4 kDa with a 14% ± 0.1%, of abundance, associated to a dimeric form of NS5. The other aggregates were visible using this HEPES buffer but were not measurable (Figure 14A).

A



C

[NS5] (mg/ml)	(S)	Abundance	f/fo	MW (kDa)
2	4.2	39%	1.8	103
2	6.6	15%	1.1	204
2	8.9	16%	0.9	480
0.5	4.1	74%	1.9	79
0.5	5.6	15%	1.4	128
0.5	7.4	6%	1.0	193
0.3	4.0	81%	1.9	92
0.3	5.6	9%	1.1	152
0.3	7.1	4%	1.0	217

B

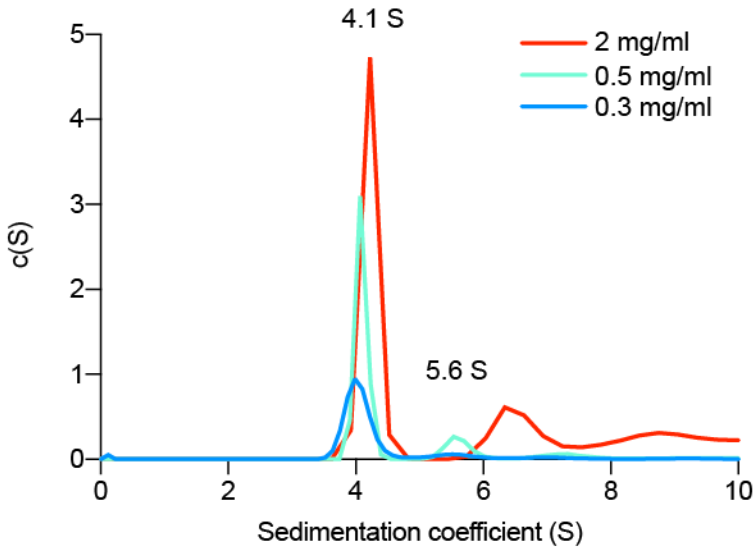


Figure 14: (A) SEC-MALS profiles of NS5 at 1.5 mg/mL in 50 mM MES pH 6.0, 500 mM NaCl, 10% Glycerol, 5 mM DTT buffer (blue solid line) and 50 mM HEPES pH 7.0, 500 mM NaCl, 10% Glycerol, 5 mM DTT (red solid line). Dashed lines represent light scattering for each condition and black traces show the measured molecular weight (right axis) of protein in the corresponding peak. **(B)** Sedimentation coefficient distribution profile of NS5 at different concentrations, indicating the values for each peak that would correspond to monomer (~4.1S), dimer (~5.6S) and oligomeric (~7.4S) species. Sedimentation coefficient with an asterisk corresponded to overlapped peaks. Sedimentation coefficients (S= are expressed in Svedberg units (1S = 10^{-13}m s^{-1}) and c(S) corresponding to normalized continuous size distribution. **(C)** Peak analysis of the sedimentation coefficient distribution of NS5. The abundance is expressed in terms of integrated signal in the corresponding s value range, relative to the total signal. Frictional ratio (f/fo) describes the sphericity of the particle, and can take values of 1.2 (perfect sphere) or higher to elongated particles.

The presence of NS5 dimers was also assayed by sedimentation velocity in analytical ultracentrifugation (AUC) using three protein concentration in buffer 50 mM HEPES pH 7.5, 500 mM NaCl, 10% Glycerol, 5 mM DTT (Figure 14B and C). Best curves were obtained using 0.5 mg/ml NS5 concentration and three species were identified in this experiment: a predominant single specie (74%) exhibiting a sedimentation coefficient of 4.1S and a frictional ratio (f/f_0) of 1.9, these results are compatible with the presence of a ellipsoidal protein with a radius of 8.96 nm that would correspond to the NS5 monomer, a second specie with a 15% of abundance in the sample, a sedimentation coefficient of 5.6S and a f/f_0 value of 1.4 with a radius of 12.92 nm could explain the presence of a dimer population. A broad population of NS5 with a sedimentation coefficient of ~7.12S was also detected (Figure 14 B). The curve analysis revealed that NS5 behaves mainly as a monomer at low protein concentration (0.5 - 0.3 mg/ml) but as the concentration raised, larger population of dimers and other larger species started to form at expenses of the monomeric population. Calculated molecular weight for the monomer and the dimer were 79 kDa and 128 kDa which differs greatly to those expected, however, this calculation was done assuming that all species presented the same hydrodynamic shape and is also influenced by the presence of several species, thus values should be interpreted with caution. Comparison of the frictional ratios showed that the monomeric NS5 adopted an elongated conformation (f/f_o ~1.9) in contrast to the dimeric forms that exhibited a more globular shape (f/f_o 1.1 to 1.4) (Figure 14 C).

3.1.5. The NS5 monomer-dimer equilibrium in solution studied by SAXS

Small Angle X-Ray Scattering (SAXS) was used to further evaluate the NS5 dimer formation in solution. SAXS is a biophysical method to study the overall shape and structural transitions of biological macromolecules, it can provide low resolution information on the shape, conformation and assembly state of proteins. This technique allows to measure the maximum particle diameter (D_{max}) and molecular weight (MW) from NS5 samples at concentrations ranging from 0.5 to 6 mg/ml (Figure 15 A). We found that SAXS calculated MW with lowest protein concentration (0.5 mg/ml) was consistent with the theoretical MW of the NS5 monomer.

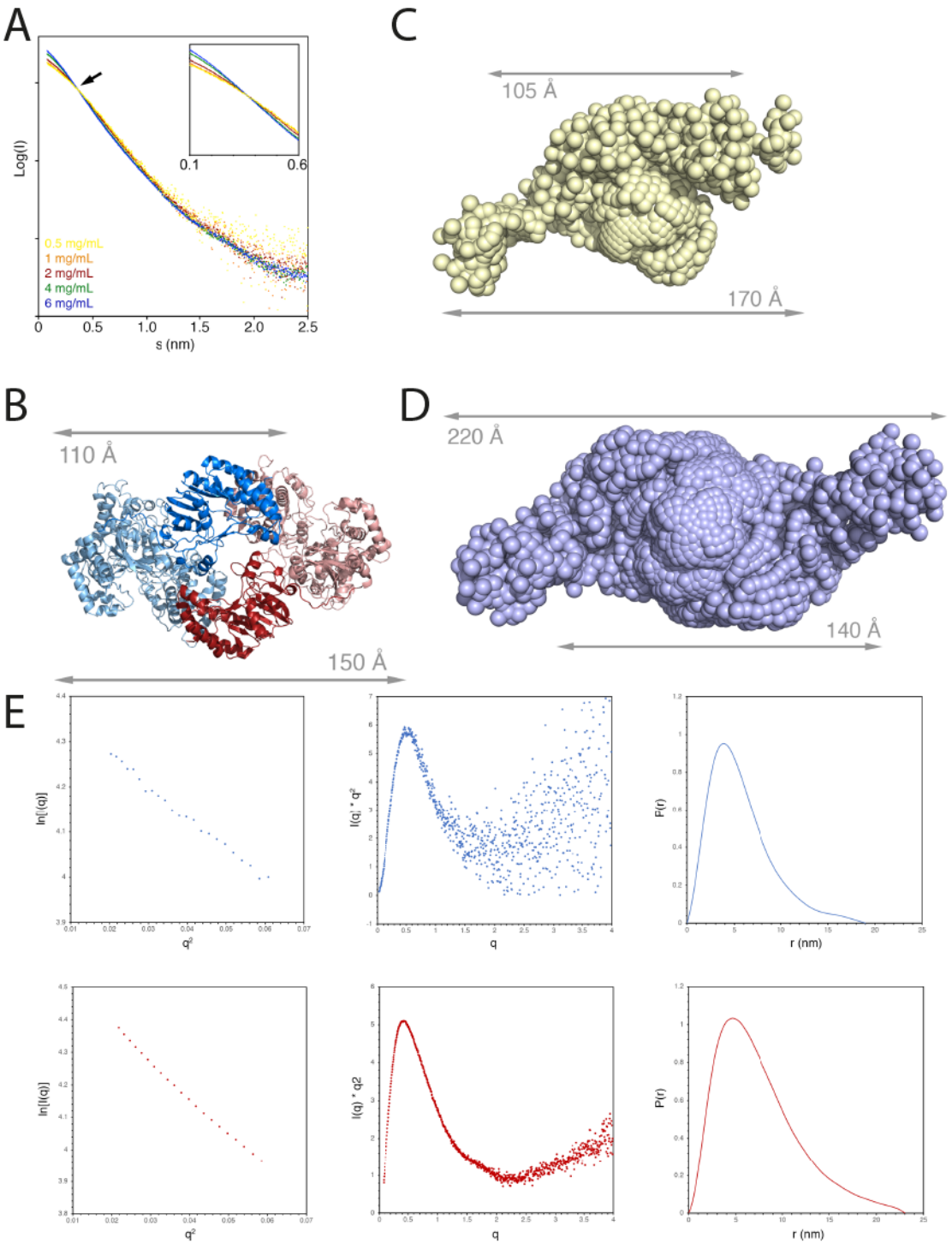


Figure 15: (A) Scattering curves at increasing concentrations of NS5. The curves collected at low (0.5 mg/ml) and high (6 mg/ml) concentrations of NS5 cross each other (black arrow) indicating the presence of different scattering species. The inset shows a detailed view of the small angle region of the curves. **(B)** Ribbon representation of the NS5 dimer, monomers are colored in blue and red, MTase domains are highlighted. Dimensions of the dimer as well as the dimer are indicated. **(C-D)** *Ab-initio* ball models generated with Gasbor corresponding to the scattering curves collected at 0,5 mg/mL (in yellow) **(C)** and 6 mg/mL (in blue) **(D)**. **(E)** Guinier regions (left panels), Kratky plots (central panels) and Pair-Distance Distribution functions (right panels) for the NS5 samples at 0.5 mg/mL (top panels, in blue) and 6 mg/ml (bottom panels, in red).

However, as the protein concentration was raised to 6 mg/mL, the calculated MW of NS5 also increased reaching 200 kDa, confirming the existence of a NS5 monomer-dimer equilibrium that is mediated by the protein concentration (Figure 15 A). We noted that scattering curves at increasing protein concentration intersected, which is indicative of a titration between two species. In SAXS, Guinier plots informs about the dimension of the particle, deviations from linearity informs about the aggregation state of the protein. Kratky plots describe the morphology of the particle and inform about the compactness of the protein. Both Guinier and Kratky plots can indicate the aggregation state of the sample. Guinier plots for all NS5 sample concentrations were of excellent quality, confirming that the increase in molecular weight was not due to protein aggregation. Similarly, the Kratky plots indicated that all samples were properly folded (Figure 15 E).

Given the good curve quality, we proceeded to calculate ab-initio models for the lowest and the higher protein concentrations (which would correspond to the NS5 monomer and dimer, respectively) (Figure 15 C and D). Calculated models exhibited excellent fits to the scattering curves with chi^2 values of 0.98 and 4.4 respectively, nevertheless the size of these models were larger than the monomer and dimers seen in crystals.

SAXS pair-distance distribution function is related to the distances within the particle, and inform about the shape of the protein. SAXS measurements at the lowest and highest protein concentration resulted in bell-shaped pair-distance distribution functions with a trailing tail. At 0.5 mg/ml, the bell shaped distribution extends from 0 to 120 Å with the tail extending to 170 Å. This distribution indicates that only few interatomic distances are that long, in turn indicating that the monomer may have a flexible region that extends beyond the folded domain.

The model derived from the crystal structure includes residues from 6 to 888 of NS5 and has 26 amino acids disordered at the C-terminus (five from the protein C-terminus plus the His-tag and the spacer). The disordered C-terminal tail is readily modeled in an extended conformation on the EOM models, reconciling the differences between the dimensions obtained from the crystal structure and the scattering curves. Similarly, the pair-distance distribution function from the sample at 6 mg/ml (at which concentration NS5 exists as a dimer), the bell-

shaped distribution extends from 0 to 170 Å with the tail extending to ~220Å (Figure 15 D). The elongated dimer observed in solution is probably caused by a combination of the volume occupied by the disordered residues in the crystal structure and small distortions of the crystallographic dimer when devoid the crystal packing constrains.

The 10-amino acid linker connecting the MTase and RdRP domains of NS5 would explain the conformational flexibility of this protein as previously seen in DENV (Bussetta *et al.* 2012, Saw *et al.* 2015, 2017 and 2019) and ZIKV NS5 (Saw *et al.* 2019). To explore whether this links also allows certain degree of flexibility of ZIKV NS5, we analyzed the scattering data for the lowest NS5 concentration (0.5 mg/mL) using EOM. Results of this analysis revealed that in spite of this linker is mobile (chi^2 of 2.4), is not completely flexible because the EOM sub-ensemble that best described the low concentration data was populated by only two conformations. However, both conformations were more extended that the crystallographic monomer indicating that, similarly to DENV, ZIKV NS5 also adopts an extended conformation in solution. This extended conformation of the model was retained at all concentrations analyzed, confirming that the NS5 dimer may retain the conformational mobility of the monomer.

3.1.6 Cross-linking coupled Mass Spectrometry experiments confirm the essential role of MTase-MTase contacts, driving NS5 dimerization

To further validate the MTase-MTase interaction seen in the crystals structures as the drivers of dimer formation, the full length NS5 as well as the individual MTase and RdRP domains were subjected to cross-linking with the label-free cross-linker suberic acid bis (3-sulfo-N-hydroxysuccinimide ester) sodium salt (BS3) (Figure 16 A). The analysis of the cross-linked complexes by SDS-PAGE showed that full length NS5 and the MTase domain are able to form dimers, as well as, higher-order oligomers in solution, whereas the RdRP domain is monomeric.

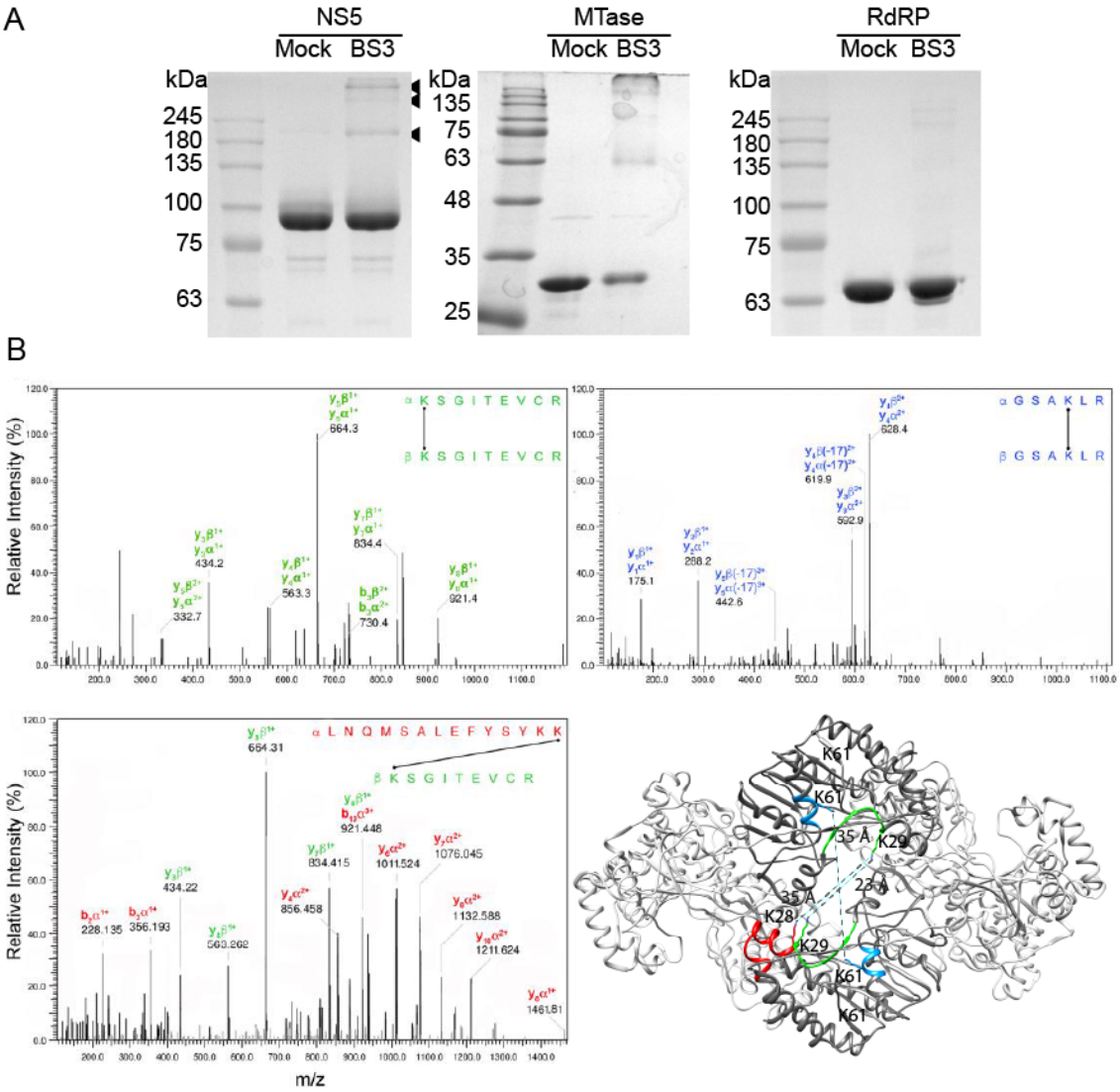


Figure 16: Cross-linking coupled to mass spectrometry (MS) analysis. Assignmet of cross-links between peptides containing K28 (red), K29 (green and K26 (cyan). The right bottom panel shows the position of the crosslinked peptides in the crystal structure of the NS5 dimer. Distances between crosslinked residues are labeled explicitly.

Crosslinked NS5 was further subjected to denaturation, trypsin digestion, and Mass Spectrometry (MS) analysis. A total of six crosslinked peptides were identified, three of them were unequivocally assigned to intermolecular crosslinks, confirming the presence of a crosslinked NS5 homodimer arranged in complete agreement to that found in the crystallographic structures (Figure 16B B). Two BS3 crosslinks involved K29 from one monomer and K28 and K29 from the second monomer and the third included a self-interaction mediated by K61 (Figure 16 B). However, in the dimer structures seen in crystals these residues are located far apart (K28-K29: 20Å; K29-K29: 22 Å and K61-K61 30 Å) to be cross-linked by BS3, which is 11.4 Å long but a rotation of the MTase domains one to each other would facilitate the joining of the involved lysines."

The other three BS3 crosslinks involved residues K458 and S452 within the RdRP ring finger. The other two pair of crosslinked residues involved peptides on both MTase and RdRP domains: a pair of crosslinked residues involving K105 within β3-η1 and T868 within α33 helix and another pair of crosslinked residues involving K96 within β6 and K843 in α31.

As shown in the NS5 structure the major MTase-MTase, interfaces, comprised residues A21, L22 and Y25 within α2 helix and K28 and K29 within α2-β1 loop, which are involved in polar and hydrophobic intermolecular contact with the α3-helix residues from K45 to V48 (Figure 17A). To confirm the role of α2 residues as the main NS5 dimerization surface, we generated two mutants on a NS5 construct with a C-terminal deletion (residues 1-886), Y25A, K28S, K29A, and A21D, L22S, Y25S, K28A, disrupting this contacting area and assayed their dimerization capacity by BS3 crosslinking and SDS-PAGE analysis. As shown in Figure 17B, mutations A21D, L22S, Y25S, K28A did not compromise dimerization since a band with a molecular weight close to dimer (180 kDa) was observed. In contrast the NS5 Y25A/K28S/K29A disrupted dimer and oligomer formation, confirming the crucial role of this contact interface and in particular, of residue K29 for NS5 dimerization (Figure 17B).

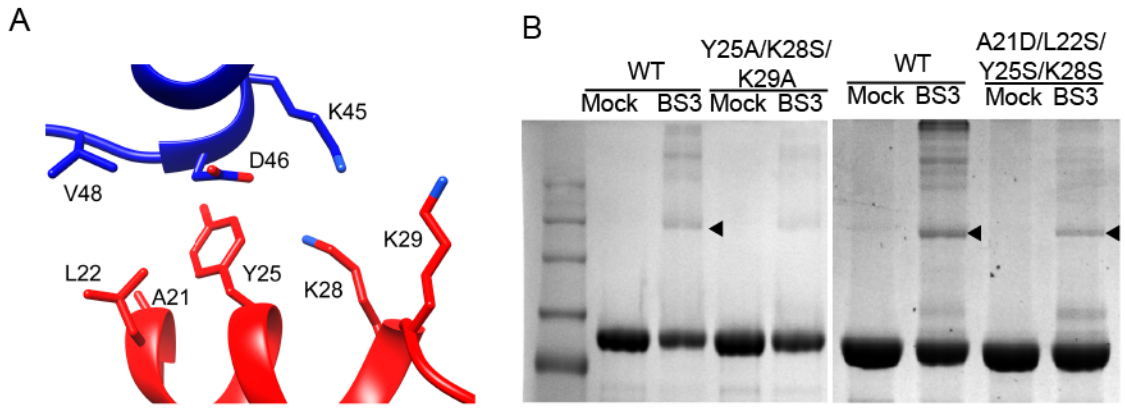


Figure 17: (A) SDS-PAGE of crosslinked NS5 WT, and its separated domains MTase and RdRP. The presence of dimers and larger species can be seen in gels corresponding to full-lenght NS5 and in the MTase domain, while the RdRP domain does not produces dimers. **(B)** Detail of the main MTase-MTase interactions, involving the α2 helix residues A21, L22, Y25, K28 and K29 of one protomer (red) with helix α3 residues K45, D46 and V48 of the second protomer (blue). **(C)** SDS-PAGE of crosslinked NS5 WT, Y25S/K28S/K29A and A21/L22S/Y25A/K28S mutant proteins. The dimer formation is indicated in lanes with an arrowhead.

3.1.7 NS5 is also able to form fibers in solution

As revealed by the crystal packing in both space groups, the dimer-based helicoidal structures involves extensive intermolecular contact areas (>1000Å^2) suggesting that these assemblies might be stable in solution and possibly inside host cells playing a role in viral replication. To asses this behavior, we imaged ZIKV NS5 protein and the NS5 Y25A/K28S/K29A mutant, with impaired dimerization activity, by negative staining EM under different experimental conditions. Experiments included low and high ionic strength buffers (ranging from 50 mM to 250 mM NaCl) and different protein concentration. Results confirmed the presence of NS5 wild-type derived fibril-like oligomers in all conditions while these high-order structures were absent when the dimerization mutant NS5-Y25A/K28S/K29A was used (Figure 18).

The presence of fiber-like structures was further assessed using atomic force microscopy (AFM) in liquid environment under the same buffer conditions (50 mM MES pH 6.0, 500 mM NaCl, 10% Glycerol, 5 mM DTT). As shown in Figure 19, AFM imaging also revealed the presence of large structures absorbed to the mica similar to that observed by negative staining EM (Figure 18 A).

These fibril-like structures showed different lengths, ranging from 20 to 300 nm length and 15 to 25 nm width (Figure 19). Although the transversal analysis of fibrils is influenced by the AFM tip (cantilever) size employed to determine the topology of the surface, driving to overestimation of the real size, the estimated values are not far from those measured in the X-ray structure of the type I dimers (Figure 19 B-D). In addition, fibrils showed 2.5 to 3.5 nm height, compatible with the dimension of a NS5 dimer laterally deposited onto mica surface and subjected to the tip compression. All these dimensions, are compatible with fibril-like structure identified in crystals packing (Figure 13) deposited on the mica surface.

As expected, the fibril-like structures are absent when the NS5 mutant Y25A/K28S/K29A mutant was imaged in the same conditions as the wild type protein (Figure 18 D and Figure 19 A).

Additional studies were performed with NS5 wild-type at different pH conditions (ranging from pH 6.0 to 8.0), showing that the fibril-like arrangement was pH

dependent. The formation of high-order structures was diminished at higher pH values, showing mainly single particles distributed on the mica surface (Figure 19 A).

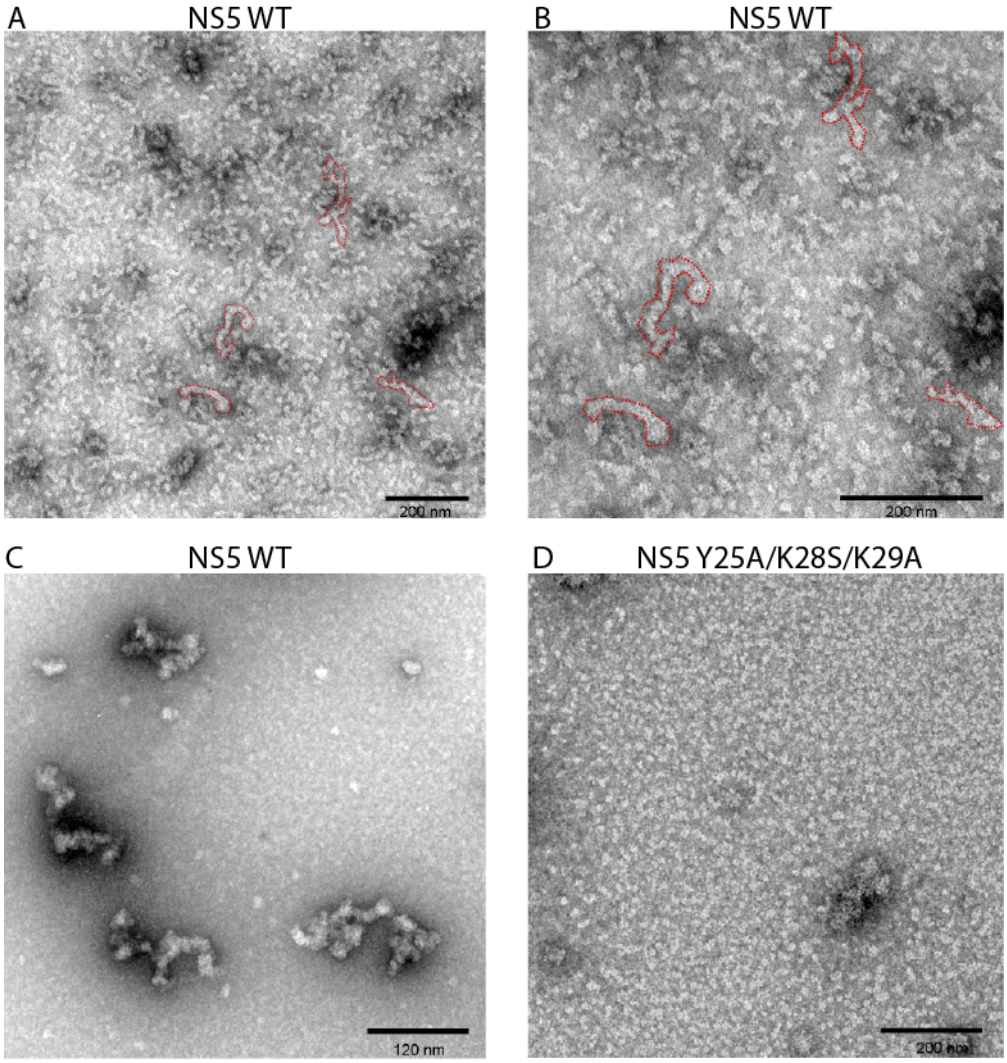


Figure 18: Negative staining electron microscopy analysis of the ZIKV NS5. **(A)** Negative-staining micrograph using 2% Uranyl-acetate of wild type NS5 protein in buffer 50 mM MES pH 6.0, 250 mM NaCl, 5 mM DTT. High order structures are highlighted in red. **(B)** Magnified area of **(A)** micrograph. **(C)** Micrograph of 1/200 dilution of **(A)** sample in the same buffer conditions. **(D)** Negative staining micrograph using the NS5 Y25A/K28S/K29A mutant at the same concentration and buffer conditions as in **(A)**.

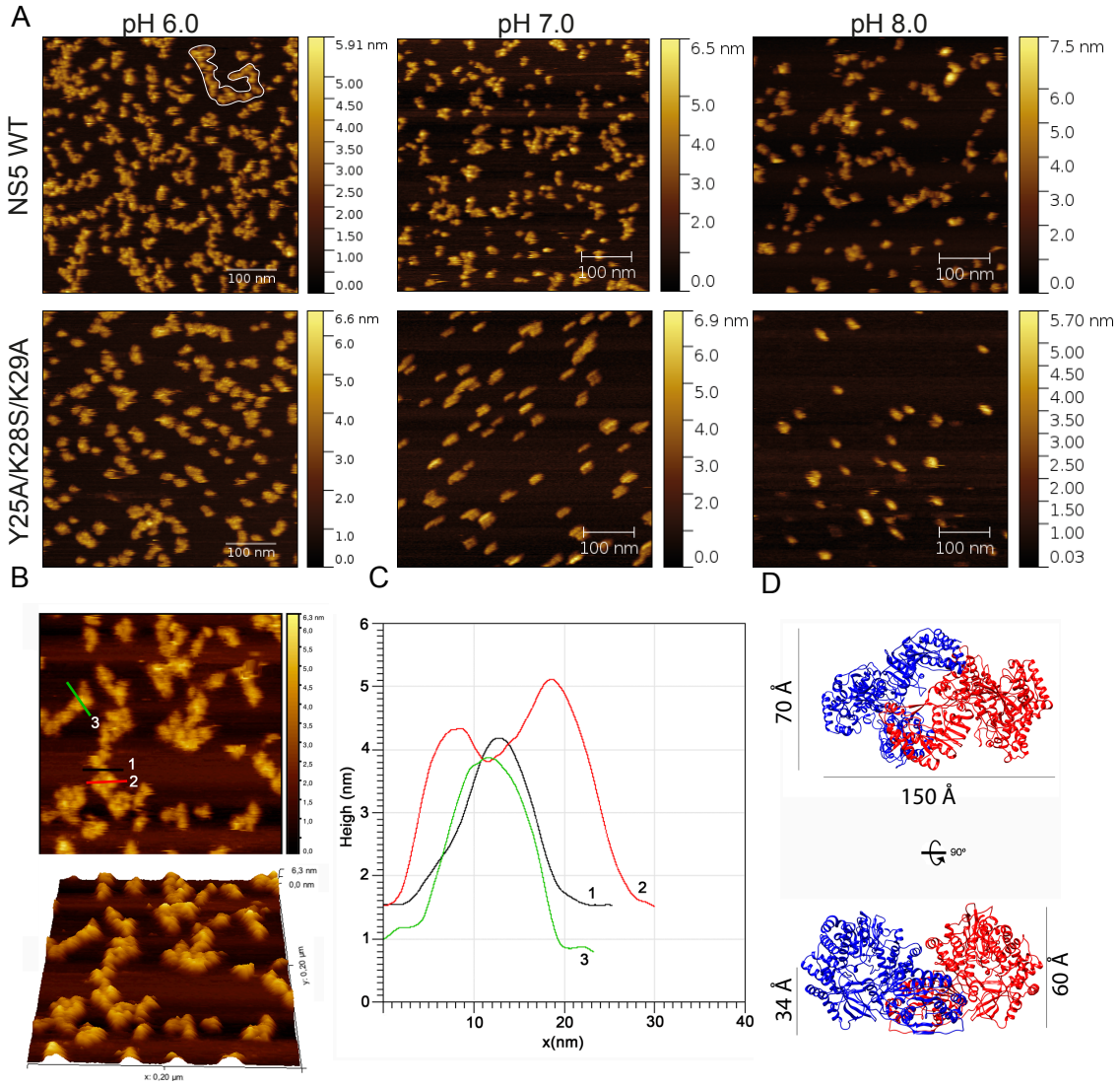


Figure 19: (A) AFM images of WT (top panel) and the mutant Y25A/K28S/K29A (bottom panel) NS5 protein at different pH values: 50 mM MES pH 6.0 (left panels), 50 mM HEPES pH 7.0 (center panels) or 50 mM Tris pH 8.0 (right panels) in 500 mM NaCl, 10% Glycerol, 5 mM DTT. Fibril-like structures can be appreciate in top-left panel encircled with a white line. High order structure formation is disrupted in the case of the Y25A/K28S/K29A NS5 mutant and results also impaired at higher pH values. Color code bar at left correspond to the measured height. **(B)** Close up image of NS5 WT in 50 mM MES pH 6.0, 500 mM NaCl, 10% Glyreol 5 mM DTT. Colored bars indicates the section areas used for measurements in **(C). (C)** Profile sections showing the dimensions of height and width of three random zones measured in the AFM image shown in panel **(B)**. Particle measurements exhibited heights and widths ranging from 3 to 3,5 nm and 15 to 20 nm respectively. **(D)** Ribbon diagram of the NS5 dimer shown in two different views, protomers are colored in blue and red respectively. Different dimensions of the dimer height and width are indicated.

3.1.8 GTP impairs NS5 dimerization

As previously described in ZIKV NS5 crystallographic structure, MTase domains were driving the dimerization by the residues Y25, K28 and K29 and by point mutation within these residues, the dimerization was impaired. For DENV, the

conserved basic residue K28 plays a major role in GTP binding as it was co-crystalized with the NS5 MTase domain (Zhao *et al.* 2015). Also, the covalent binding of this K28 residue to GTP as an intermediate state during guanylyltransferase (GTase) activity of Flavivirus NS5 was already described (Issur *et al.* 2009). As both dimerization and GTP binding activities involved the residue K28, we assessed whether the GTP impaired the dimerization using BS3 as crosslinking agent in presence and absence of this nucleotide. To further confirm that the interaction was GTP specific, we also performed the experiment in presence of ATP as negative control and in presence of GMP. These experiment revealed that ATP did not impair the dimer formation while the dimer formation was altered when NS5 was incubated in presence of GMP or GTP showing almost no preference for these two nucleotides (Figure 20). These results suggest that dimerization might act as a negative regulator of the GTase activity of NS5 which is a crucial step in the capping process.

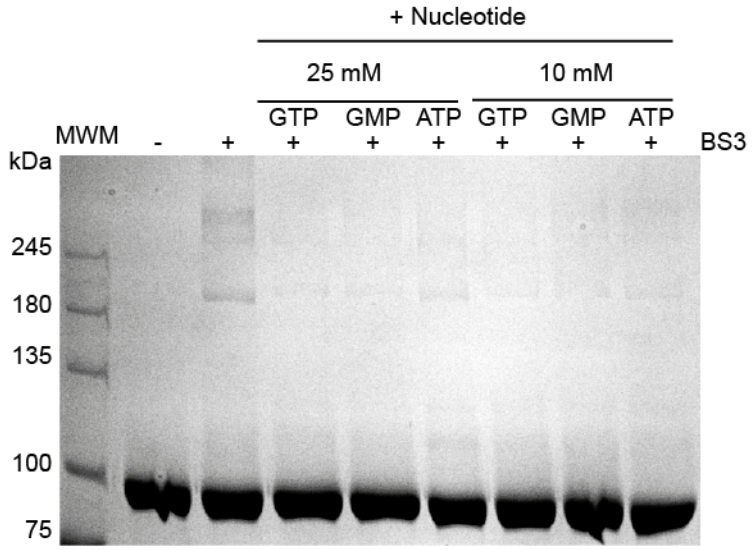


Figure 20: SDS-PAGE of cross-linked NS5 in absence or in presence of the nucleotides: GTP, GMP or ATP at 25 mM and 10 mM nucleotide concentration. The dimer formation remains unaltered in the case of ATP while in the case of both GTP and GMP the dimerization is impaired.

3.1.9 In vitro RdRP activity of recombinant ZIKV NS5 proteins

In-vitro RNA synthesis activities of the purified wild type NS5 and dimerization mutant (Y25A/K28S/K29A) were measured in a primer extension assay with a fluorescent labeled ssRNA oligonucleotide (20-mer) previously annealed to an unlabeled ssRNA oligonucleotide (28-mer) which yielded a 3’ overhang that

would be extended by the NS5 polymerase. As result of NS5 complete RNA polymerization, a band shift from 20-mer to 28-mer was observed in denaturing gel electrophoresis. To compare the elongation activity of wild-type and dimerization mutant ZIKV NS5 protein, a parallel assay using several protein dilutions were used. As shown in Figure 21, the NS5 Y25A/K28S/K29A mutant was about three times more active than the wild type NS5 at higher protein concentration where quaternary structure is promoted, suggesting that NS5 dimerization down-regulates the RdRP elongation activity.

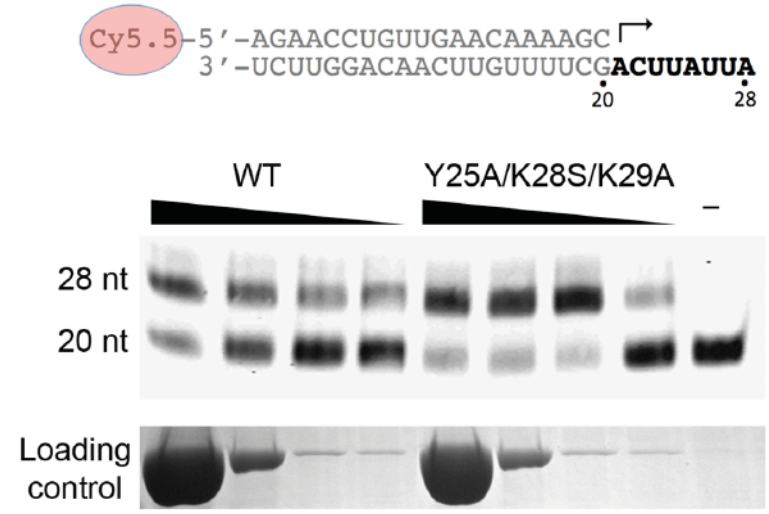


Figure 21: (A) Polymerase activity evaluated in a nonradioactive primer extension assay. The top panel shows the template/primer duplex used in the assay. The arrow indicates the location and direction of primer extension. The bottom panel shows the effect of the wild type NS5 or NS5-Y25A/K28S/K29A dimerization mutant concentrations (900, 180, 18, 9 nM) on polymerization activity. This assay used a constant template/primer concentration of 10 nM.

3.1.10 Study of NS5 *in vivo*

The embryonic structure that originate the central nervous system in all vertebrates is the neural tube (Greene *et al.* 2009) that primarily consist of neural progenitor cells (NPC), a population of proliferative cells with multipotent differentiation ability (Martínez-Cerdeño *et al*. 2018). Previous studies showed that ZIKV infection dysregulates proliferation and differentiation of NPC (Gabriel *et al*. 2017, Yoon *et al*. 2017). In collaboration with Elisa Martí and Murielle Saade, we evaluated the effect of ZIKV NS5 protein *in vivo*, using a neural tube of chick embryo as model.

The expression of ZIKV NS5-GFP fusion protein by electroporating half part of the chick neural tube was successful as showed in green in Figure 20A. A considerable part of the viral protein exhibited a nuclear localization as expected due to the presence of NLS signals (Brooks *et al*. 2002), showed as yellow color by the overlapping of red color channel (nuclear staining) and green (NS5-GFP) (Figure 22 A). Surprisingly, NS5 expression produce an abnormal growth of the neural tube in comparison to GFP as negative control, suggesting that NS5 was involved in the phenotype of viral infection (Figure 22 A).

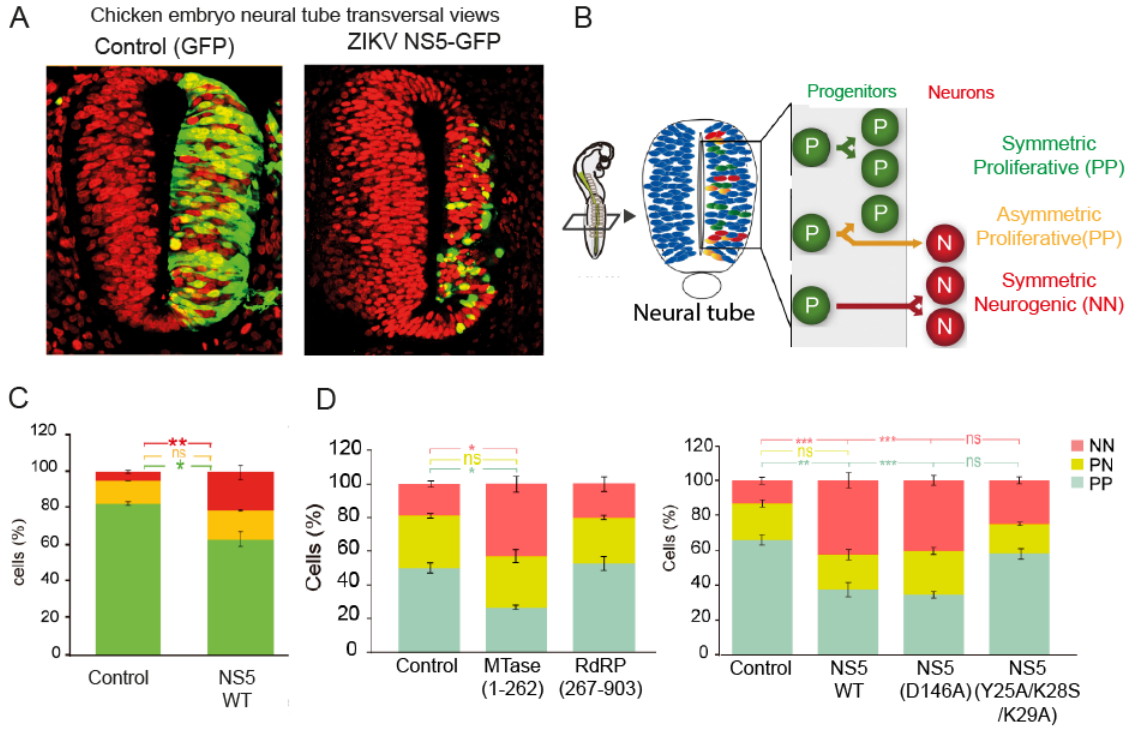


Figure 22: *In*-vivo effect of ZIKV NS5 on the neural tube of chicken embryo. **(A)** Neural tube of chicken embryos. Left side of the neural tube was not electroporated as control and righ side was electroporated with GFP (left panel) or NS5-GFP (right panel). **(B)** Schematic representation of the three modes of division of NPC. **(C) - (D)** Quantification of reporter-expressing FACS sorted cells in each condition: PP ($pSox2^+/pTis21^-$, green), PN ($pSox2^+/pTis21^+$, yellow) and NN ($pSox2^-/pTis21^+$, red). The data represents the means ± s.e.m.)

Normal growth of the neural tube requires a fine balance between cell proliferation and cell differentiation. This is carried out by the coexistence of three types of NPC division: symmetric proliferative divisions (PP) produces two NPC daughter cells, asymmetric divisions (PN) that produce one NPC and one neuron and symmetric neurogenic divisions (NN) that produce two neurons (Figure 22 B).

These modes of division can be individually tracked by the expression of the reporter genes fused to fluorescent proteins: Sox2-EGFP (PP), Sox2-EGFP and

Tis21-RFP (PN) and Tis21-RFP (NN) generating a color-coded indicative (green, yellow and red respectively) of division mode (Saade *et al.* 2013). To assess whether ZIKV NS5 affects the mode of NPC division, chick embryos were co-electroporated with plasmids for NS5 together with the pSox2-EGFP/pTis21-RFP reporter genes. When the NPC cells of control (not electroporated) were analyzed through fluorescent assisted cell sorting (FACS), the majority of cells were under PP mode of division, as expected for a neural tube under development. However, a decreased rate of PP was observed in NS5 electroporated samples at the expense of an increase in neurogenic divisions (NN) (Figure 22 C). To evaluate the role of the individual domains and their specific activities (MTase or RdRP) in the modes of division, the same experiment was performed with plasmids expressing the isolated domains. The expression of MTase was able to reproduce the results of full length NS5, producing a significant increase in the NN mode of division in contraposition to RdRP domain (Figure 22 D). Results with a mutant of MTase active site (D146A) demonstrate that the activity of this domain was not responsible for the observed phenotype (Figure 22 D), producing similar increase in NN mode of division to the WT domain. Interestingly, mutant Y25A/K28S/K29A did not significantly altered the mode of division in comparison to the control sample. These results, indicate that residues involved in ZIKV NS5 dimerization are required to alter the mode of division promoting neurogenesis and prematurely consuming the NPC population.

3.2 Interactions of NS5 with the 5SLA genome promoter

3.2.1 5SLA RNA in-vitro transcription and purification

To obtain structural information of the interactions between ZIKV NS5 and the RNA promoter 5SLA, we synthesized and tested different constructs derived from the ZIKV 5'UTR sequence (Table 2). They vary in the length of the stem and two of them included three final GC clamp. Finally, a 5SLA sequence construct with three terminal GC clamp (construction 5SLA GC) was selected based on a previously published study of the ZIKV genome folding (Ziv *et al.* 2018) (Table 2) (Figure 23 A and B) and construct stability.

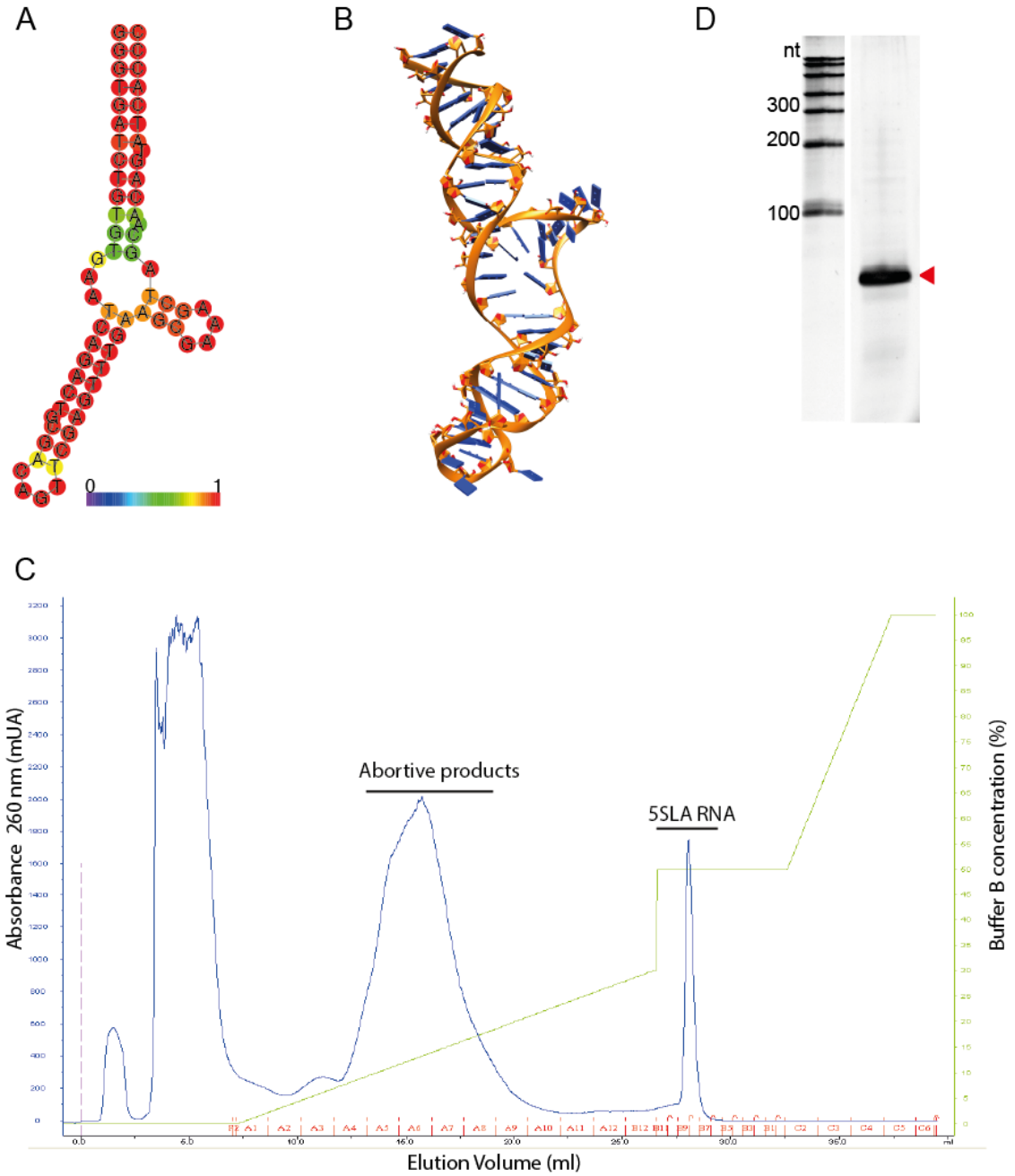
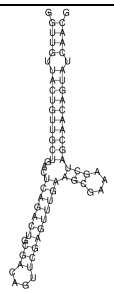
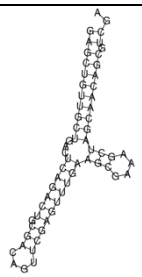
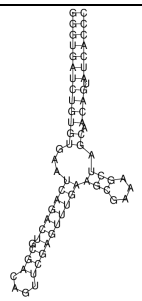


Figure 23: (A) Schematic representation of the 5SLA RNA base-pair interactions. The structure is colored by base-pairing probabilities as indicate bottom bar. **(B)** Predicted 3D model for the 5SLA sequence. E total value was -1409. **(C)** Chromatogram of a purification of 5SLA RNA, abortive transcription products as well as the 5SLA RNA peaks are explicitly indicated. **(D)** Denaturing 8% PAGE showing a 5SLA sample (red arrow) after concentration of the 5SLA RNA peak from (C).

In-vitro transcription conditions for large scale RNA production were optimized. Several variables as template DNA, nucleotide, DTT, T7 polymerase concentrations as well as transcription time were optimized and reactions were analyzed on denaturing Urea PAGE and quantified. Best conditions were found using 50 ng/µL DNA, 15 mM DTT, 0.5 mg/mL T7 polymerase 5 mM NTP and used in successive RNA productions. A 10 ml reaction approximately yielded

~150 µg of RNA. Transcribed RNA was purified using an anion exchange chromatography in a standard FPLC system, elution profiles exhibited a single peak at absorbance at 260 nm that corresponded to the ZIKV 5SLA RNA construct (Figure 23 C). Fractions corresponding to this peak were analyzed by denaturing Urea PAGE and pooled together confirming the presence of one band of the expected size and appropriate homogeneity for biochemical and structural studies (Figure 23 D).

Table 2: ZIKV 5SLA constructs tested. Construct name, length, sequence as well as predicted folding are indicated in the table. Prediction of the base pairing interaction was obtained from RNAfold Server (Gruber *et al.* 2008).

Construct name	Length (nt)	Sequence (5'-3')	Folding
5SLA Long	70	GGUUGUUACUGUUGCUGACUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACG	
5SLA Short	63	GAGCUGUUGCUGACUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUAGCAACAGCGUCGA	
5SLA GC	67	GGGUGAUCUGUGUGAAUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUAGCAACAGUAUCACCC	
5SLA Motif	46	GGGUGACUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUACCC	

3.2.2 Electrophoretic Mobility Shift Assay reveals that NS5 binds the 5SLA construct

Electrophoretic mobility shift assay (EMSA) was performed to asses the NS5 binding capacity for 5SLA RNA. As the gel running conditions are not denaturing, the forces that stabilize the complex remain along the run, maintaining the native intermolecular interactions. Serial dilutions of NS5 and constant concentration of the 5SLA (0.45 µM) covering several molar ratios, were tested to find the amount of protein necessary to produce the RNA band shift. Gels were stained with SYBR Gold and visualized using UV. The shift was clearly obtained using NS5:5SLA 1:1 molar ratio (Figure 24 A). The RNA shifted band exhibited increased intensity as the protein concentration increased, in agreement with concomitantly formation of new NS5-5SLA complex. However, higher protein concentrations in the reaction led to protein stacking to the gel well, this could be due to protein instability at high concentrations in the EMSA buffer.

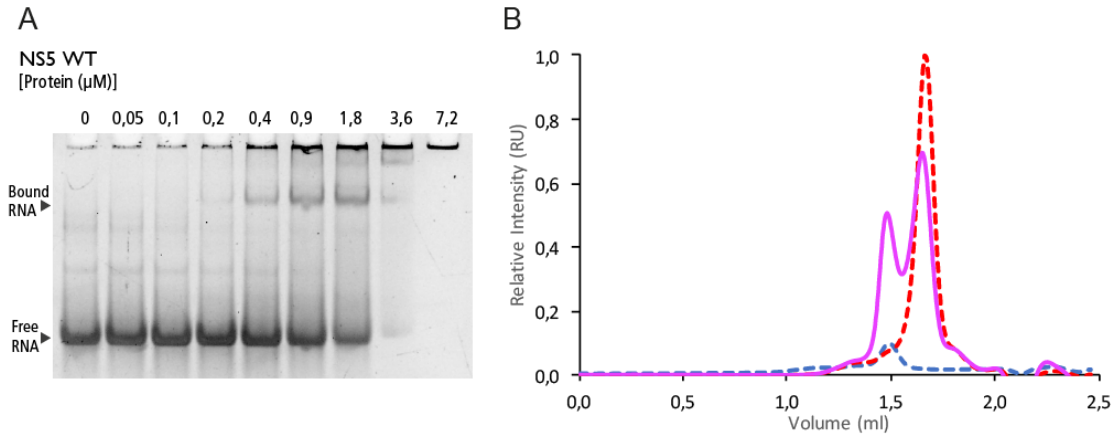


Figure 24: (A) Electrophoretic mobility shift assay (EMSA) results showing the binding of NS5 and 5SLA. Protein concentrations are listed above each lane. The RNA concentration was maintained constant at 0,45 µM in all lanes. RNA was stained with SYBR Gold. **(B)** Size exclusion chromatogram showing the NS5-5SLA binding. Absorbance values were registered at 260 nm, relativized to the 5SLA run and depicted in dashed blue line for NS5, in dashed red line for 5SLA and in solid purple line for the NS5 5SLA complex.

3.2.3 Size Exclusion Chromatography confirms the stability of the NS5-5SLA complex

Despite the fact that the EMSA is employed to positively evaluate protein-RNA complex formation, some nucleoprotein complexes do not resist the physical forces occurring in SEC experiments, quickly disassembling and making it difficult

to analyze with cryo-EM. To assess the strength of the ZIKV NS5-5SLA complex, we subjected it to an analytical SEC experiment (Figure 24 B). To avoid dilution effects and to increase the resolution capacity, we used a Superdex 200 3.2/30 gel filtration column, previously equilibrated with the stabilization buffer (50 mM Bis-Tris Propane pH 8.0, 300 mM KCl, 10 mM DTT).

Elution volumes as well as absorbance ratio 260/280 nm ($Abs_{260/280}$) data was used to follow the complex formation. Separate runs of pure NS5 and 5SLA RNA samples with elution volumes of 1.5 ml ($Abs_{260/280}$: 0.5) and 1.6 ml ($Abs_{260/280}$: 2.33) respectively, were used as references. The NS5-5SLA complex elution profile exhibited two peaks at 1.5 and 1.6 ml, overlapping previous NS5 and 5SLA reference peaks. However, the intensity of the peak at 1.6 ml that corresponded to the 5SLA, showed decreased absorbance at 260 nm in comparison to the reference run. On the other hand, the peak that eluted at 1.5 ml ($Abs_{260/280}$:1.12) that corresponded to the NS5 showed increased absorbance at 260 nm if compared with the NS5 reference run (Figure 24 B). The increased absorbance confirmed that NS5 bound 5SLA strongly and the intermolecular interactions resisted the physical forces that occurred in the size exclusion column, thus being a good cantidate for cryo-EM experiments. However, as NS5 and the NS5-5SLA elution volumes were indistinguishable, we could not isolate a homogeneous complex.

3.2.4 Optimization of NS5-5SLA sample for cryo-EM

To evaluate the NS5-5SLA complex quality and concentration we used negative staining electron microscopy (NS-EM). The first NS5-5SLA complexes prepared in the stabilization buffer [50 mM HEPES pH 7.5, 250 mM KCl, 1 mM $MgCl_2$, 5 mM DTT], exhibited non-homogeneous particle distribution allover the grid support. Also, particle morphology was not uniform, indicating that further optimization of buffer composition was required to improve the sample quality (Figure 25).

Nucleic acid-protein complexes fall apart in presence of high salt concentrations, so ionic strength should ideally remain as low as possible, but high enough to

keep NS5 soluble. To assess whether the complex was folded at low salt concentrations, we performed Dynamic light scattering (DLS), EMSA and Differential scanning fluorimetry (DSF) experiments to determine the stability and binding capacity in presence of different salt concentration

The aggregation state of the complex was checked by DLS using salt concentrations ranging from 100 to 400 mM KCl and NS5 5SLA samples were diluted to 10 μM in the appropriate buffer before measuring. Data corresponding to intensity and volume size distribution was used for evaluation (Figure 25 A and B). The size distribution by intensity is a direct representation of the DLS measurement, this detects very sensitively large particles as they scattered much more light than small ones as the intensity of scattering of a particle is proportional to the sixth power of its diameter. By contraposition, size distribution by volume informs about the volume distribution or abundance of the particles and is calculated from the intensity. Results showed a strong NS5-5SLA aggregation at 100 mM KCl as peaks corresponding to particle diameter of 1.000 to 10.000 nm appeared in both plots. Different large species with diameters ranging from 100 to 10.000 nm were also present when the buffer contained 400 mM KCl though this peak was not present in the volume plot probably because the low population. However, when 250 - 300 mM KCl was used, peaks corresponding to a single specie with a diameter of ~10 nm appeared at both intensity and volume plots. This size was compatible with the expected size of the NS5-5SLA complex. However, a slight shift in diameter particle using 250 mM KCl toward larger values than expected, lead us to set the salt concentration to 300 mM KCl (Figure 25A).

To assess the capacity of NS5 to bind 5SLA RNA in this range of salt conditions, the binding affinity was evaluated by EMSA in presence of increasing KCl concentrations, ranging from 100 mM to 400 mM in the stabilization buffer. EMSA revealed a clear shifted RNA band that appeared at NS5 concentrations from 0.22 to 3.6 μM in all buffers (Figure 25 C). Under the light of DLS and EMSA results and considering that high salt concentrations such as 400 mM KCl are not appropriate for cryo-EM studies, we set the salt concentration 300 mM KCl for the complex formation

To further optimize buffer composition, an extensive buffer screening using Differential Scanning Fluorimetry (DSF) was performed as previously described (see M&M section, Chari *et al* 2015). Using this method, the stability of the NS5-5SLA complex was analyzed using 8 different buffers at 8 different pH conditions, resulting in a total of 64 screened conditions. Fluorescence emission was recorded from 4 ºC to 95 ºC and first derivative was calculated for each sample and plotted for each buffer compound (see Figure 1 in Annexes). The buffer Bis-Tris methane pH 7.2 exhibited the best unfolding curve, raising the melting temperature value up to 42 ºC, first derivative fluorescence curve at this pH also exhibited a narrow shape indicating a fast protein unfolding without partial denaturalization (Figure 25 D). As Bis-Tris methane presented better melting temperatures and curve shapes at higher pH values, we decided to increase the pH to 8.0, however, as Bis-Tris methane does not buffer properly at this pH, we used Bis-Tris propane instead. To evaluate whether magnesium played a role in NS5-5SLA complex formation and stabilization, EMSA assays were performed in presence and absence of Mg^{2+}. As showed in Figure 25 E, no significant differences were detected when the divalent ion was present, and hence, Mg^{2+} was discarded from the stabilization buffer. After all optimizations were done, the final buffer composition was set to 50 mM Bis-Tris Propane pH 8.0, 300 mM KCl and 10 mM DTT.

To evaluate the sample quality in the new buffer condition, NS5-5SLA complex was imaged using negative staining transmission electron microscopy (NS-TEM). We observed a homogeneous particle dispersion allover the carbon grid support, with regular particle morphology, size and without aggregates (Figure 25 F and G).

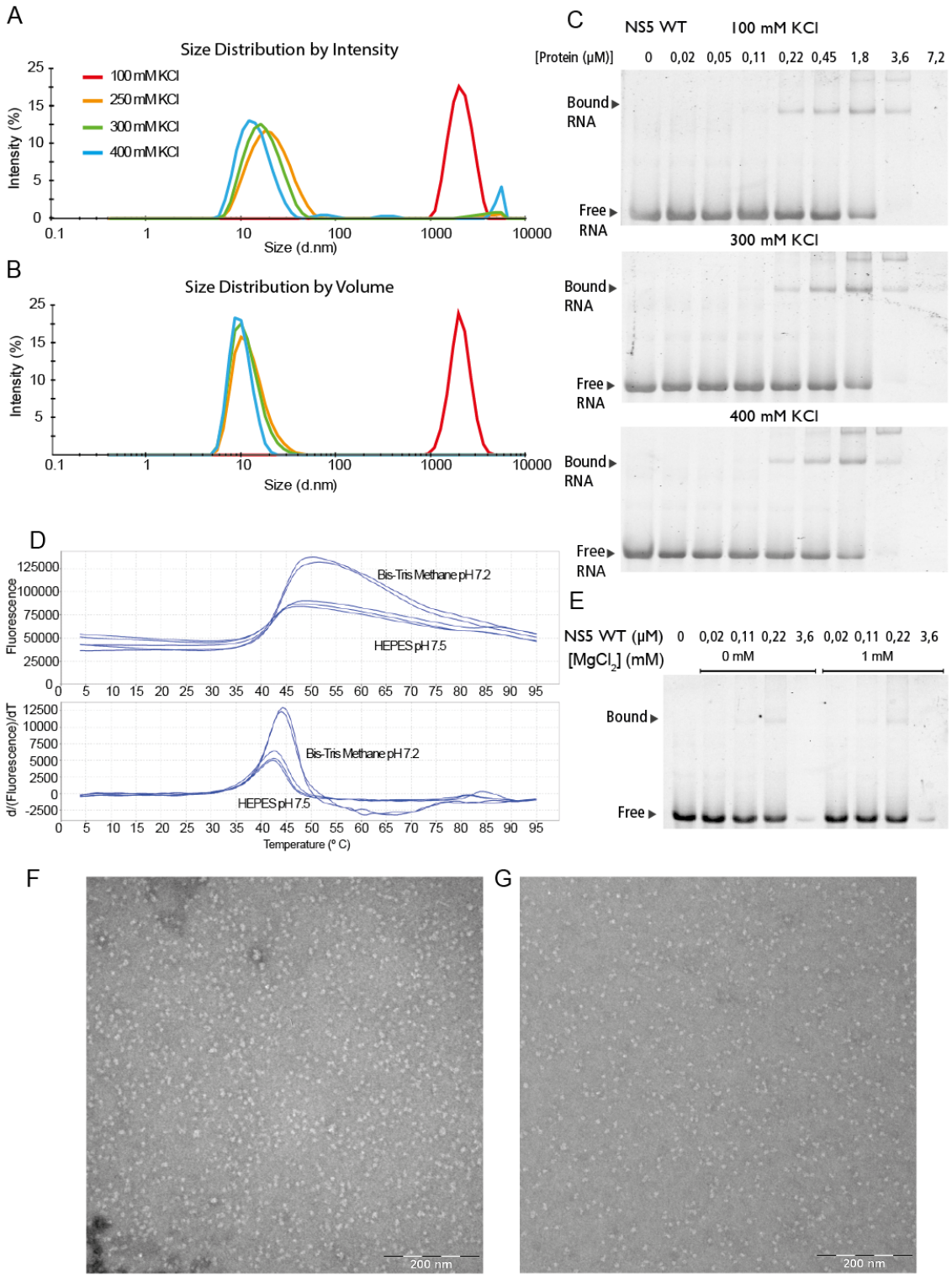


Figure 25: Sample optimization of the ZIKV NS5-5SLA complex. Dynamic Light Scattering (DLS) data of the NS5 5SLA complex in buffer 50 mM HEPES pH 7.5 at different KCl concentrations. Plots correspond to **(A)** size distributions by intensity and (**B**) size distribution by volume. **(C)** EMSA data showing the binding of NS5 and 5SLA at different KCl concentrations. Protein concentrations are listed above each lane. **(D)** Differential Scanning Fluorimetry (DSF) results of the NS5 5SLA complex assayed in HEPES pH 7.5 (three replicates) and Bis-Tris Methane pH 7.2 (two replicates). Curves correspond to the direct fluorescence readings (top) and first derivative of the fluorescence (bottom). **(E)** EMSA data showing the binding of NS5 and 5SLA in absence (left lanes) and presence (right lanes) of 1 mM of $MgCl_2$. NS5 protein concentratios are listed above each lane. Negative Staining micrographs of the NS5 5SLA complex **(F)** in the

original buffer 50 mM HEPES pH 7.5, 250 mM KCl, 1 mM MgCl2, 5 mM DTT and **(G)** the optimized buffer 50 mM Bis-Tris Propane pH 8.0, 300 mM KCl, 10 mM DTT.

3.2.5 Cryo-EM studies

Sample preparation for NS-TEM differs greatly from that for cryo-EM. Unlike NS-TEM, in cryo-EM, samples are embedded in amorphous ice within the grid holes. This implies that even though the particle concentration was optimized using NS-TEM as described in the Materials and Methods section, it was necessary to re-optimize the particle concentration using cryo-EM. Several complex concentrations, ranging from 50 µM to 5 µM were assayed. The best particle distribution all over the grid holes was found at 5 µM of NS5-5SLA complex (Figure 26). Two types of grids were also assayed: C-Flat multihole and 2/2 Cu 300 mesh and Quantifoil 1.2/1.3 Cu/Rh 300 mesh, these last ones were the best for cryo-EM imaging (Figure 26).

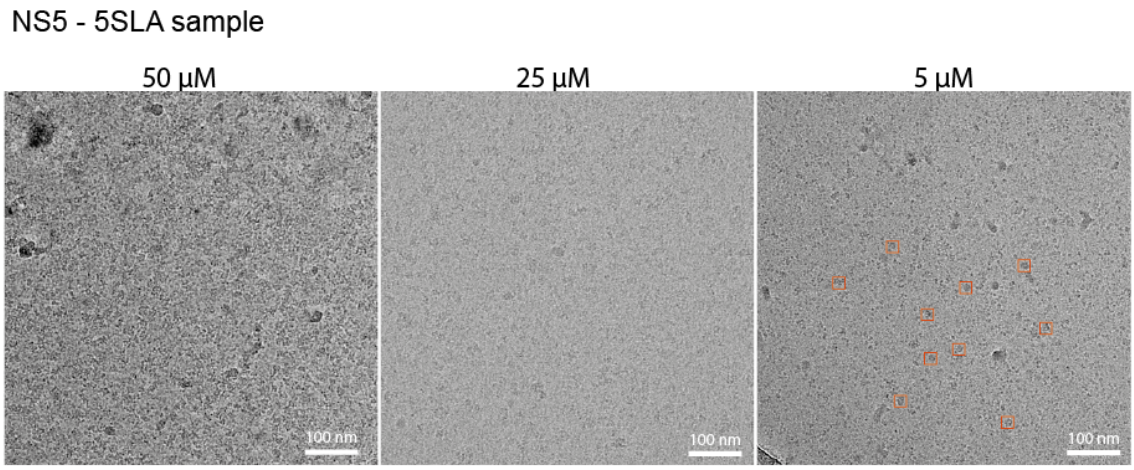


Figure 26: cryo-EM micrographs of different sample concentrations of NS5-5SLA complex. Samples at 50 and 25 µM were vitrified using C-Flat 2/2 Cu 300 grids, while the sample at 5 µM was vitrified using Quantifoil 1.2/1.3 Cu/Rh 300 grids. Isolated particles are indicated with red boxes in the right micrograph. Micrographs were acquired using a 200 kV Tecnai G2 microscope with a cryo transfer holder and a CCD detector using the low the dose mode.

After optimizing the NS5-5SLA complex concentration, several cryo-EM grids were vitrified and checked using a 200kV Talos Arctica microscope, equipped with a Falcon III direct electron detector. Vitrified samples showed good ice thickness gradient all over the grid that allowed the identification of best ice-thickness squares. Several test images were acquired to check particle concentration and morphology, showing a moderate-high concentration of particles in the grid holes with a strong increment in particle concentration

towards the hole edges. Due to the small particle size and the difficulty to differentiate from background, several movies were recorded and motion-corrected to confirm that the observed particles corresponded to expected size of the complex, that should be approximately 100 Å (Figure 27).

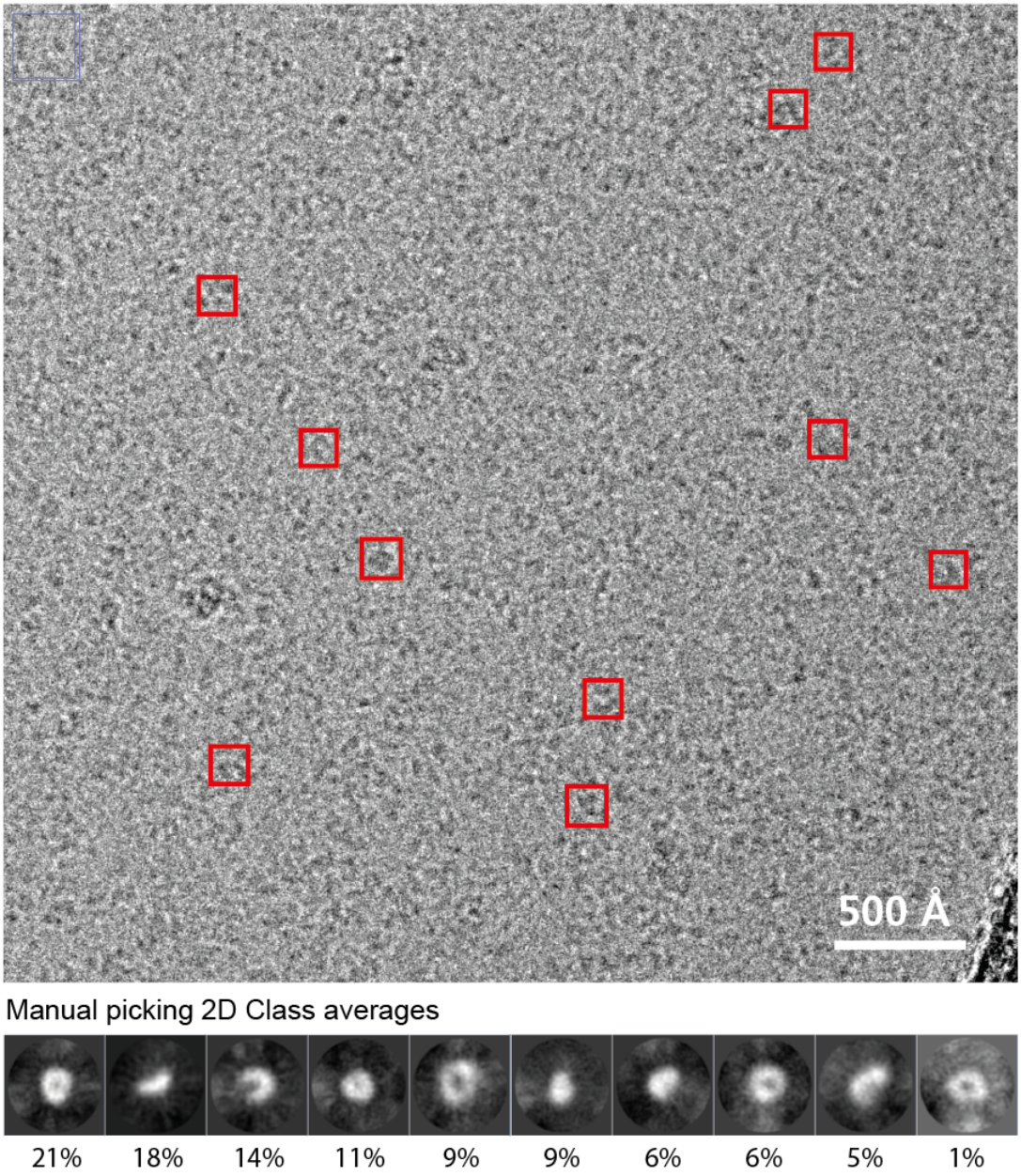


Figure 27: Representative motion corrected cryo-EM micrograph of the NS5-5SLA complex (top). NS5-5SLA sample at 3.5 µM was vitrified using Quantifoil 1.2/1.3 Cu/Rh 300 grids. Micrograph was obtained using a 200 kV Talos Arctica microscope with a Flacon III detector. Isolated particles are indicated with red boxes. 2D class averages corresponding to 3700 manually picked particles are showed at the bottom panel. Population percentage to each class is indicated under each class average.

A total of 808 movies of 64 frames each were collected at different defocus values (Table 3). Movies were motion-corrected using gCTF (Zhang *et al.* 2016), two

sets of a total of 3.770 and 15.000 particles were manually picked from dose-weighted micrographs at different defocus values and submitted to a reference free 2D Classification with RELION 3.0 (Scheres *et al.* 2012). 2D-Class averages obtained with these manually picked particle sets did not exhibited differences that may be caused by the number of particles, both 2D averages showed a “ring-shaped” and “half moon” forms (Figure 27). These manually obtained 2D classes were used as input templates for autopicking using Gautomatch (Zhang *et al.* 2011), which resulted in an automatically picked 202.382 particle dataset. A reference free 2D Classification performed with this dataset, revealed the presence of six 2D classes that exhibited different particle sizes: one subset of three classes of smaller particles (158K particles) of ~65 Å length and another subset of three classes of larger particles (42K particles) of ~95 Å (Figure 28). Both subset of particles were used to reconstruct an initial map with RELION (Figure 28). Smaller particles generated an initial map with similar size and shape to the NS5 (Figure 28), however, further 3D classification and 3D refinement did not improve the reconstruction. Obtained reconstruction using the subset of larger particles resulted in a map that could harbor a monomer of NS5 and an extra density that could be attributed to the RNA. This map was used as reference for 3D classification with 4 classes. This classification separated one class composed by 27K particles that reconstructed a map in which a NS5 monomer could be fitted, showing an extra density connecting MTase and RdRP map regions (Figure 28). As 3D Refine did not improved the map quality we performed signal subtraction in RELION by masking the NS5 map, thus eliminating the signal that corresponded to the 5SLA. This new subset was refined producing a ZIKV NS5 map at 9.4 Å of resolution (Figures 28 and 29). The low resolution obtained could be related to the small size of the protein and to the high flexibility of both protein domains. A larger data acquisition with a 300 kV microscope and better detector might increase the resolution of the reconstruction. The X-Ray crystal structure of ZIKV NS5 (PDB: 6I7P) was rigid-body fitted into the NS5 cryo-EM map and Real Space refined in Phenix.

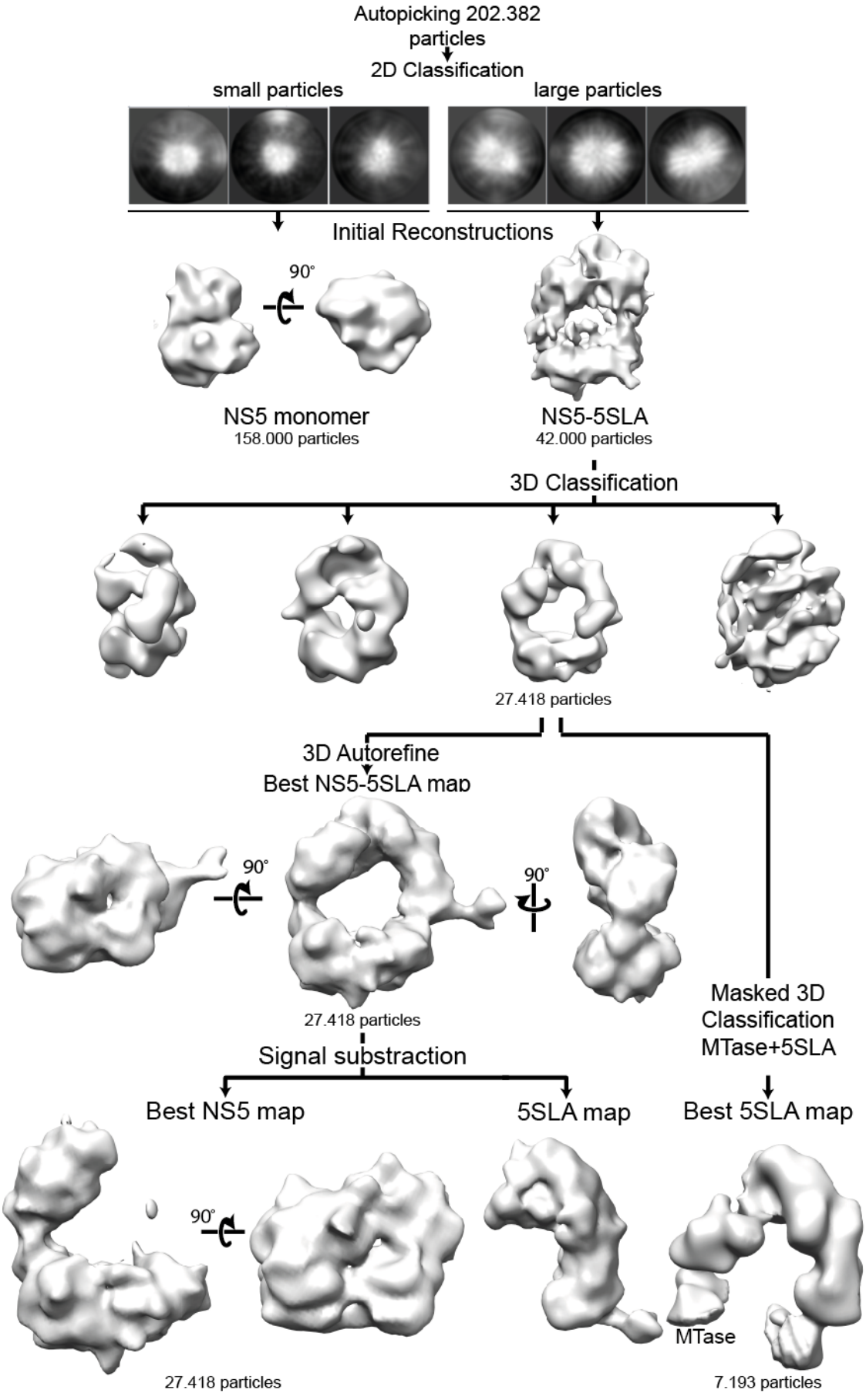

Autopicking 202.382 particles
2D Classification
small particles
large particles
Initial Reconstructions
90°
NS5 monomer
158.000 particles
NS5-5SLA
42.000 particles
3D Classification
27.418 particles
3D Autorefine
Best NS5-5SLA map
90°
90°
27.418 particles
Signal substraction
Masked 3D Classification MTase+5SLA
Best NS5 map
5SLA map
Best 5SLA map
90°
MTase
27.418 particles
7.193 particles

Figure 28: Workflow of the NS5 5SLA cryo-EM processing. 2D Class averages generated from the manually picked particles that were used as template for autopicking. 202.382 automatically picked particles were submitted to 2D classification and produced three different classes with sizes compatible with monomeric NS5 (small particles) and three classes with larger sizes compatible with the NS5-5SLA complex (large particles). We used both subsets to to build initial reconstructions of the monomeric NS5 and the NS5-5SLA complex. 3D Classification with the particles corresponding to the NS5-5SLA initial reconstruction, produced a map that was interpreted by the presence of the bound 5SLA. This subset of 27.418 particles was refined and produced the best NS5-5SLA reconstruction. The 5SLA map region was subtracted and refined yielding the refined NS5 map. The best 5SLA reconstruction was obtained using a masked 3D classification on the MTase and 5SLA map regions.

Table 3: Cryo-EM data collection of the NS5-5SLA complex.

Cryo-EM Data Collection			
Microscope	Talos Arctica		
Voltage (kV)	200		
Camera	FEI Falcon III		
Magnification	120.000		
Pixel size (Å/px)	0.85		
Total exposure (e-/Å2)	32.10		
Exposure time (s)	41.69		
Number of frames per exposure	64		
Data collection software	EPU		
Defocus range (μM)	1.0 to 2.5		
Number of micrographs	808		
Number of initial particles	202.382		
Number of particles for 3D analysis	42.707		
Symmetry	C1		
	NS5-5SLA complex	NS5	5SLA
Number of final particles	27.418	27.418	7.193
Resolution (Å)	11.12	9.37	11.8

To improve me 5SLA map quality, we repeated the signal subtraction eliminating the signal that corresponded to the NS5, however, 3D Refinement of these particles did not increased the map quality (Figure 28). As we expected higher flexibility in both MTase and RNA, we performed a masked 3D Classification with four classes on these map regions. This generated a 5SLA reconstruction were a helical shape could be appreciated in the map (Figure 28).

A 5SLA model was built *de novo* using Autodrrafter (Kappel *et al.,* 2019) using the 5SLA 3D predicted model as template and the masked 3D classification map of the 5SLA (Figure 23 B, 28 and 30 A).

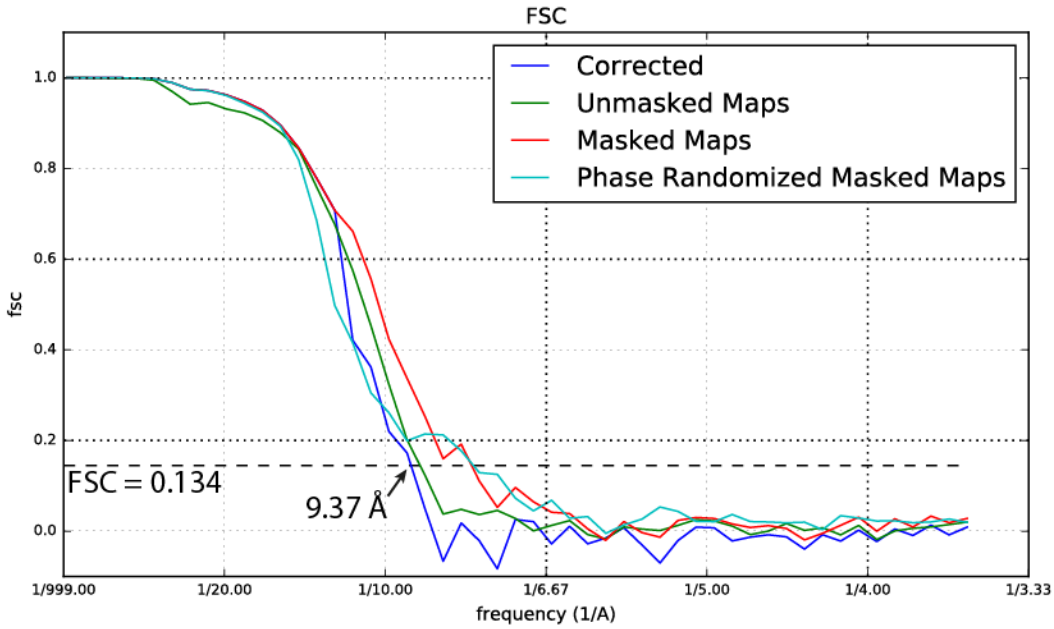


Figure 29: Fourier Shell Correlation (FSC) plot for the NS5 cryo-EM density map. Black arrow indicates the estimated resolution of the NS5 map using the FSC = 0.134 criteria.

3.2.6. The cryo-EM structure of NS5-5SLA complex

The NS5-5SLA complex shows a ring-like structure in which the NS5 molecule uses both domains, MTase and RdRP to bind the RNA molecule (Figure 30 A).

Despite the moderate resolution of the cryo-EM map, initial fitting of the X-ray coordinates of the full length NS5 into the cryo-EM maps revealed a displacement of 23 Å and a rotation of 14º in the orientation of the MTase domain relative to the RdRP, possibly as a consequence of 5SLA binding (Figure 30 C). In addition, large conformational changes in the RdRP domain were also observed (Figure 31). Due to these conformational changes, the RdRP was divided into the three independent subdomains: fingers, palm and thumb that were independently fitted into the density as rigid bodies. Also the priming loop that serves as a stabilizing platform for *de novo* initiation of RNA replication, appeared displaced 24 Å and rotated 70º from the position observed in the crystal structure, leaving the central cavity fully open (Figure 30 C, 31 and 32).

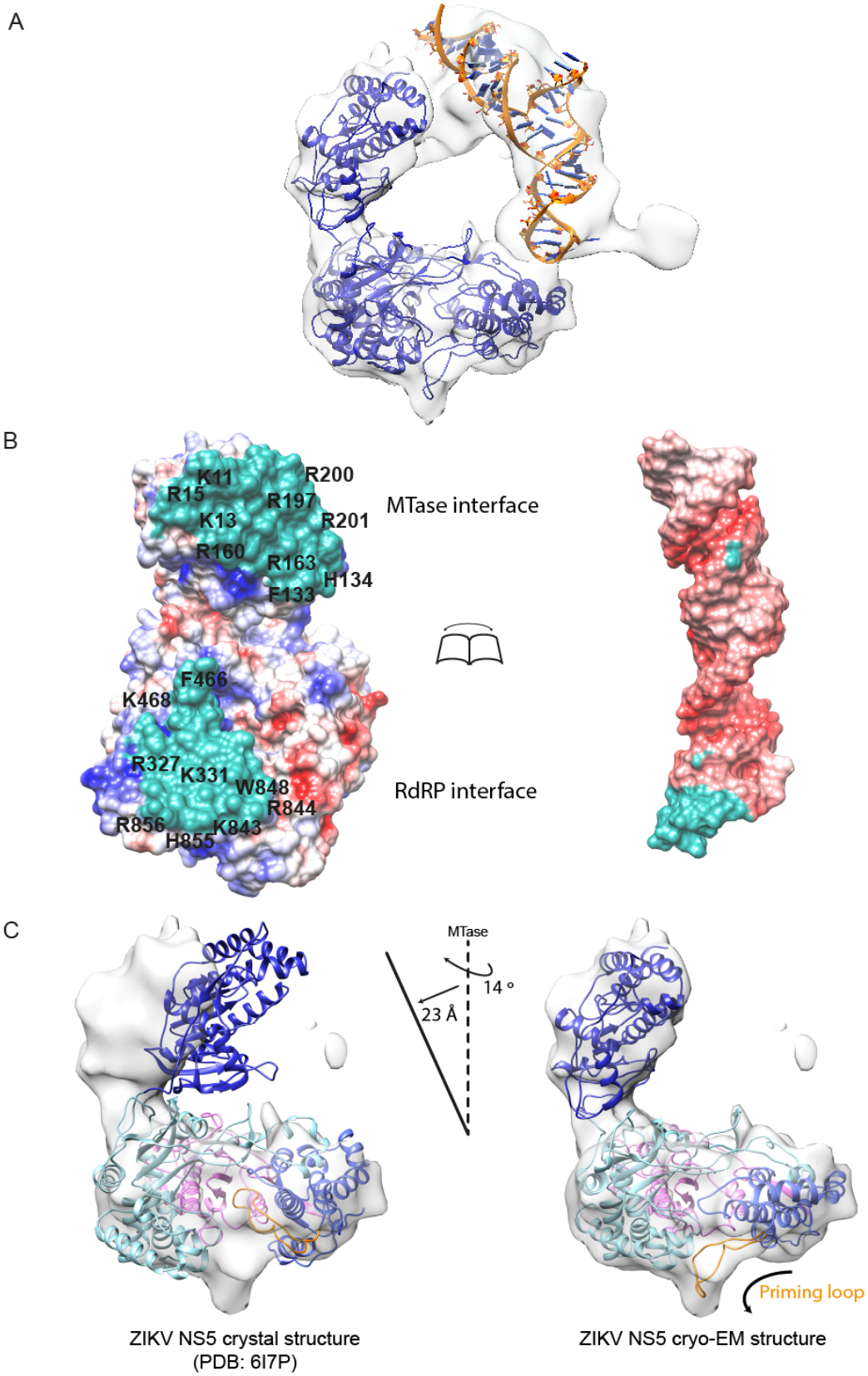


Figure 30: (A) Structure of the ZIKV NS5 (blue) in complex with the 5SLA RNA (golden) into its respective signal subtracted cryo-EM maps. Surface area of both interfaces were calculated using the NS5-5SLA cryo-EM map, and comprises ~1889 Å^2 and ~1496 Å^2 for MTase and RdRP interfaces respectively. **(B)** Coulombic surface representation of NS5 (at left) and 5SLA (at right). NS5 - 5SLA Interaction surfaces are depicted in sea green color. Potential interacting exposed residues of NS5 as well as 5SLA nucleotides are labeled in black. **(C)** Comparison of the obtained

ZIKV NS5 crystal model (PDB: 6I7P) (at left) and cryo-EM model (at right) fitted into the NS5 cryo-EM maps. MTase is colored in dark blue, RdRP sub domains are coloured in light cyan: fingers, in purple: palm and in blue: thumb. The priming loop is colored in orange. At the center, a representation of the displacement and rotation of 23 Å and 14 º between the crystal and cryo-EM MTase model.

The 5SLA in flavivirus is prediced to adopt a stem loop structure (Figure 2) (Dong *et al.* 2008, Lodeiro *et al.* 2009) however, due to the limited resolution of the reconstructed maps, neither the details of the RNA secondary structure elements nor the accurate positioning of the 5SLA 5' and 3' ends could be defined with confidence. Previous studies described that the 5SLA top loop binds to the NS5 RdRP thumb (Hodge *et al.* 2016), the MTase domain was also known to bind RNA but with negligible specificity (Yap *et al.* 2015, Hodge *et al.* 2016) and hence, the orientation of the 5SLA into the cryo-EM maps was modeled based on this information.

Two major interaction surfaces are seen in the ZIKV NS5-5SLA complex: interface A comprises the MTase domain and both 5' and 3' 5SLA RNA ends, and the interface B comprises the RdRP thumb subdomain and the top loop of the 5SLA (Figure 30 A and B).

MTase interface comprises a surface of ~1889 $Å^2$ within this domain, composed mainly by basic residues such as arginine (K11, K13, R15, F133, H134 R160, R163, R197, R200 and R201) (Figure 30 B). RdRP interface is also located within a basic patch at the thumb subdomain and comprises a surface of ~1496 Å within this domain where several basic and aromatic residues are exposed (R327, K331, F466, K468, K843, R844, W848, H855 and R856) (Figure 28 B).

As mentioned above, significant rearrangements occur in the NS5 molecule as consequence of RNA binding: i) changes in the relative disposition between the MTase and the RdRP domain (Figure 30 C) and ii) re-orientation of the thumb domain relative to the fingers and movement of the priming loop (Figure 31 and Figure 32). The movement between the MTase and RdRP domains was described before in the crystal structure of DENV 3 NS5 (Zhao *et al.* 2015). The most relevant changes occur in the RdRP thumb subdomain that undergoes large structural rearrangements.

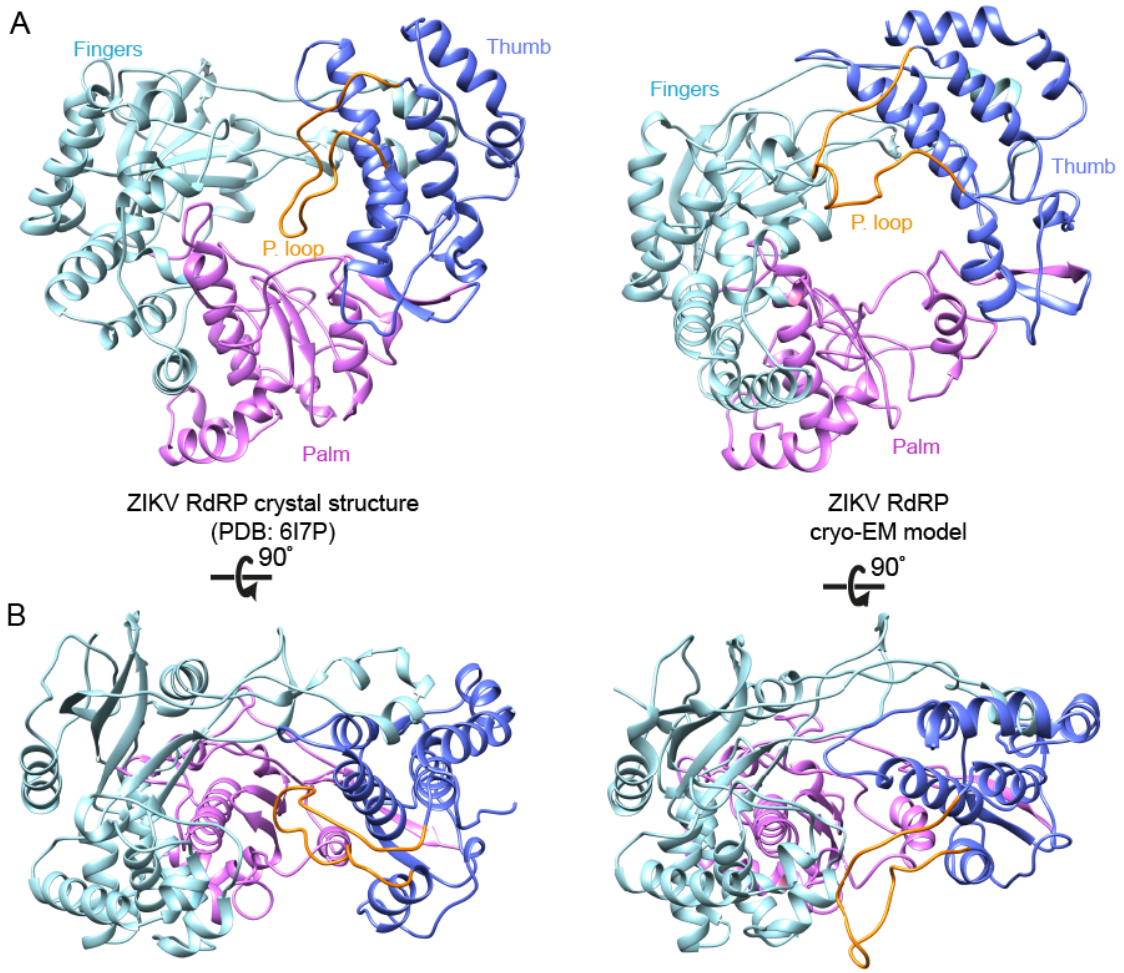


Figure 31: Structural changes in the ZIKV NS5 RdRP domain on 5SLA binding (A) Side by side comparison of ZIKV NS5 RdRP structures (front views) found in crystals of isolated NS5 (left panel) and in the cryo-EM structure of NS5-5SLA complex (right paned) models. The fingers, palm and thumb sub domains are colored in light cyan, purple and blue, respectively. The priming loop is colored in orange. **(B)** Top views of the ZIKV RdRP crystal (left) and cryo-EM (right) structures colored as in A.

In the ZIKV NS5 crystal structure, the priming loop occludes the central channel preventing the 3' end of the template ssRNA to approach the active site. However, in the cryo-EM structure, this loop appears displaced ~24 Å and rotated ~70º from its original position and with the the movement of the thumb, open the RdRP central channel (Figure 31 and Figure 32), increasing its volume up to 5.100 $Å^3$, a 3.7-fold increment compared to the unbound NS5 (Figure 33 A). Differences in charge were appreciated in the central cavity, where more basic residues get exposed to the solvent probably to ease the housing of the dsRNA (Figure 33 A and B). Other changes can also be found in the Fingers subdomain that compacts over the active site, both the fingers displace ~6 Å towards the

central channel other movements such the index tip also displaces up to 10.8 Å conducting the thumb movement and the rest of fingers (Figure 31 and 33 C).

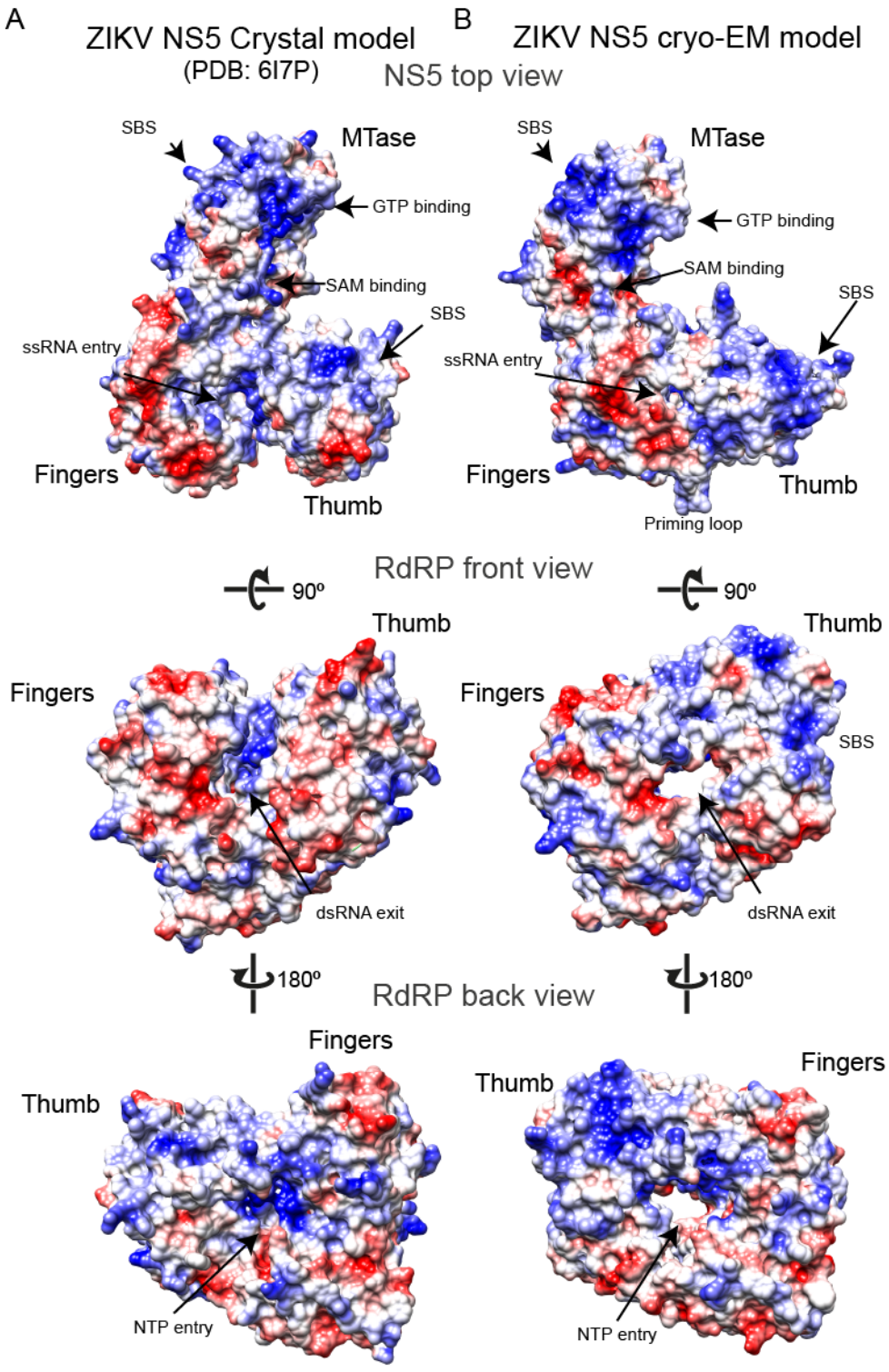


Figure 32: Electrostatic surface of ZIKV NS5 is shown in three orientations in both crystal **(A)** and cryo-EM **(B)** models. NS5 subdomains, channels and ligand binding pockets are indicated. 5SLA binding sites are indicated as SBS.

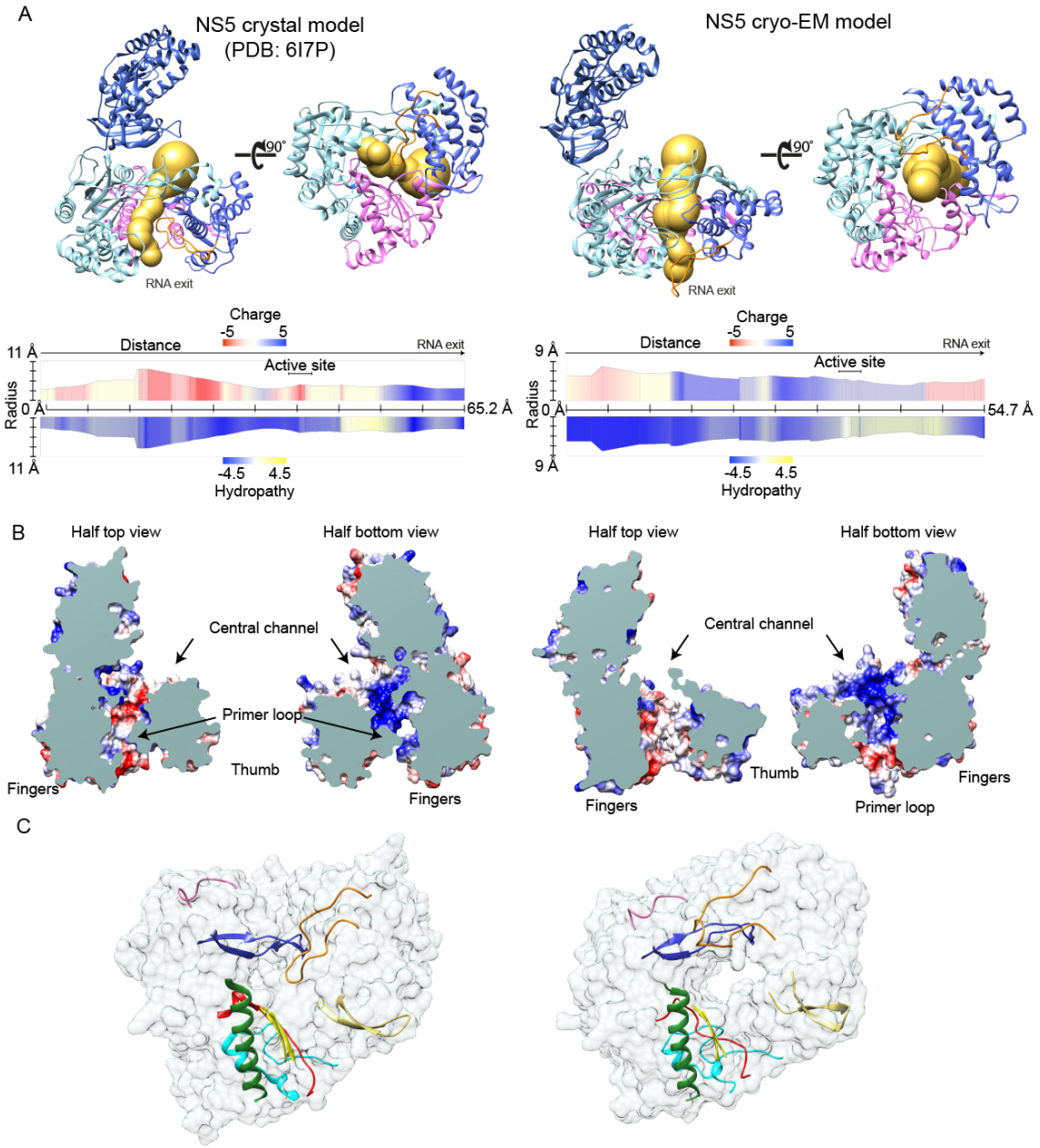


Figure 33: Analysis of the NS5 central channel on crystal (left panels) and cryoEM models. **(A):** Tunnel found of the central channel of NS5 (in golden) using Mole server (Pravda *et al.* 2018). The average charge and hydropathy of the surrounding residues on the tunnel is plotted in the bottom panel and goes from the back to the front part of the NS5 RdRP. The position of the active site within this profile is indicated. **(B):** Culombic surface representation of sliced top (left) and bottom (right) views of the NS5 central channel. The position of the RdRP domains as well as the priming loop is indicated explicitly. **(C)** Front surface view of the distribution of the conserved RdRP motifs in both crystal and cryo-EM models. RdRP conserved motifs are depicted in ribbon and are colored in red (motif A), in green (motif B), in yellow (motif C), in cyan (motif D), in wheat (motif E), in blue (motif F) an in pink (motif G). The priming loop is depicted in orange.

3.2.7 Validation of the NS5-5SLA interactions by Hydrogen Deuterium Exchange- Mass Spectrometry experiments

Protons present on macromolecules are labile and easily exchanged with hydrogen groups in solution. Hydrogen deuterium exchange mass spectrometry (HDX-MS) take advantage of this phenomena and provides solvent accessibility information that can be used to map intermolecular interactions (Deng *et al* 2016). To confirm the interactions observed in the cryo-EM structure and also uncover other transient interactions, HDX-MS experiments were carried out, using the NS5 alone, the NS5-5SLA complex and NS5 with a single stranded unstructured RNA as control of unspecific RNA binding. Samples were deuterium exchanged, and peptides that bound specifically 5SLA were detected from those that did not, as these peptides were not accessible to the solvent, and hence, not exposed to the deuterated buffer (protected peptides). Data from the NS5 with non-structural RNA was used to discern between the specific or non-specific RNA binding but only small differences were observed between these samples. Differences observed in the deuterium uptake between NS5 (and similarly to the unstructured RNA) and NS5-5SLA sample revealed several peptides that bound specifically 5SLA (Figure 34).

We found that the peptide located at the MTase, containing the residues from 222-240 that belongs to the β-8 and α-8 (Figure 10) appeared to be protected and overlaps the NS5-5SLA interface A (Figure 30 A). Several peptides were also identified in the RdRP domain, overlapping the NS5-5SLA interface B. e.g. residues from 828 to 848 that comprises the η9 and α-9 fragment or the peptide comprising the residues from 854 to 872 at the α-36. HDX experiments also revealed additional peptides that appeared to bind 5SLA through other interfaces, perhaps more transient, not seen in the cryo-EM structure.

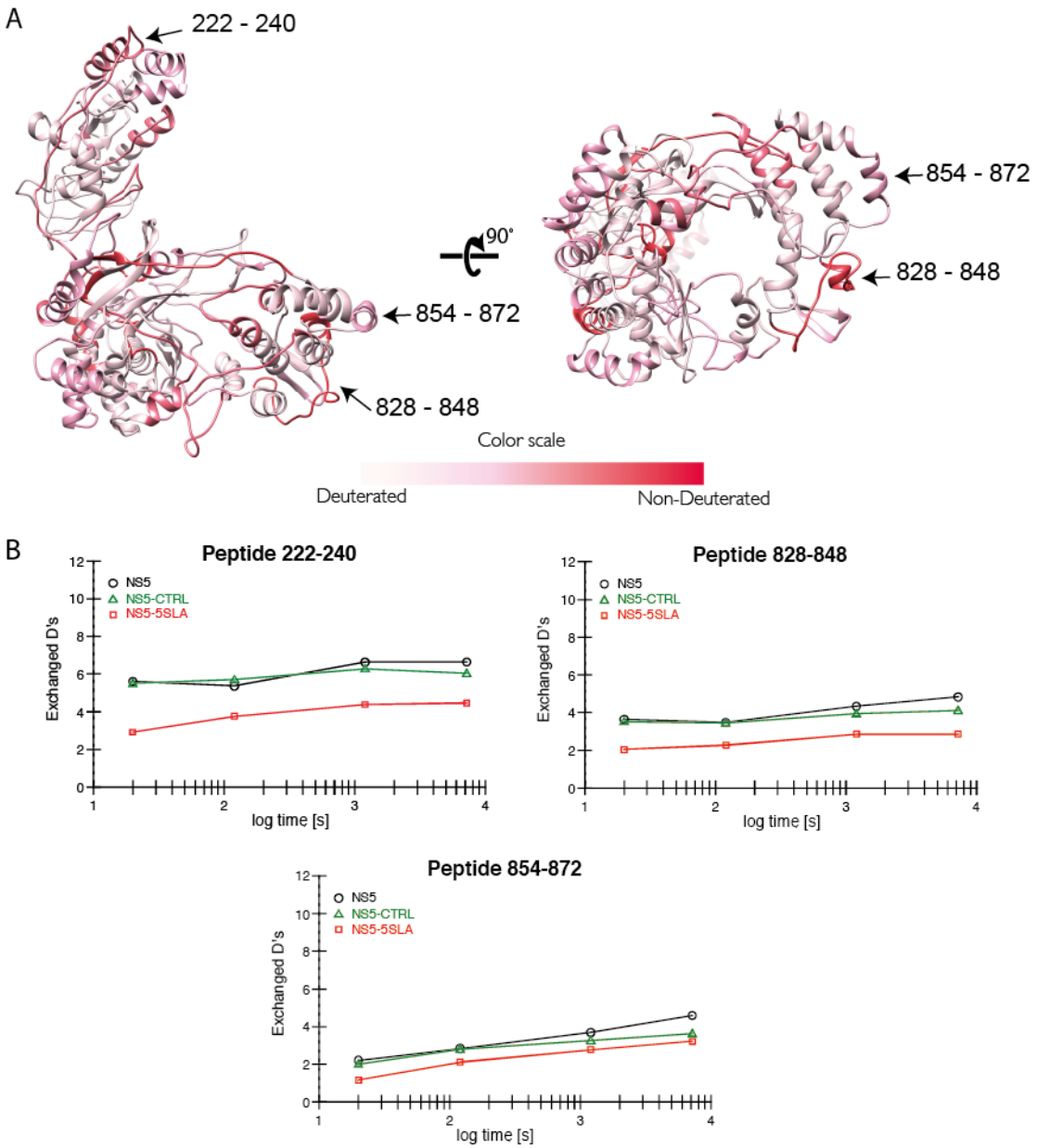


Figure 34: (A) HDX characterization of the NS5-5SLA binding. Deuterium uptake differences between the NS5-unstructured RNA and the NS5-5SLA are mapped on the NS5 structure determined by cryo-EM depicted in ribbon. Differential peptides are color-coded according to the bottom scale bar from pale-pink (peptides exposed to deuterium and not binding 5SLA) to red (5SLA binding peptides not exposed to deuterium). Peptides that overlap the 5SLA interacting surfaces in the NS5-5SLA cryo-EM maps are indicated with black arrows. **(B)** Amount incorporated deuterium within each NS5 peptide highlighted in **(A)** for each time point.

3.2.8 Biochemical validation of NS5-5SLA interfaces

To further validate the interactions observed in cryo-EM complex, new EMSA assays were performed with increasing concentrations of the individual MTase or RdRP domains of NS5 to quantify the 5SLA binding. As shown in Figure 35 A, well defined band shifts were observed when 5SLA was in presence of the MTase

domain. At the same time, decreased amounts of free RNA were observed at the gel bottom. Similarly, when the 5SLA was incubated in presence of the RdRP domain, a shifted band was present. The band corresponding to the free RNA was also diminished at high RdRP concentrations (Figure 35 A). The quantification of this EMSA revealed that the RdRP domain presented a greater affinity for the 5SLA than the MTase domain and furthermore, as seen in the binding curves, this domain exhibited higher Hill coefficient (an indicative of the stepness of the curve and cooperativity), suggesting the presence of more complex interactions within this domain (Figure 35 B). The greater affinity of the ZIKV NS5 RdRP domain for the 5SLA that we obtained is in agreement with recently published data (Bujalowki *et al.* 2020) although our calculated affinity slightly differed from those described by Bujalosky and collaborators. These differences could be attributed to the different biochemical techniques used to calculate the binding affinities.

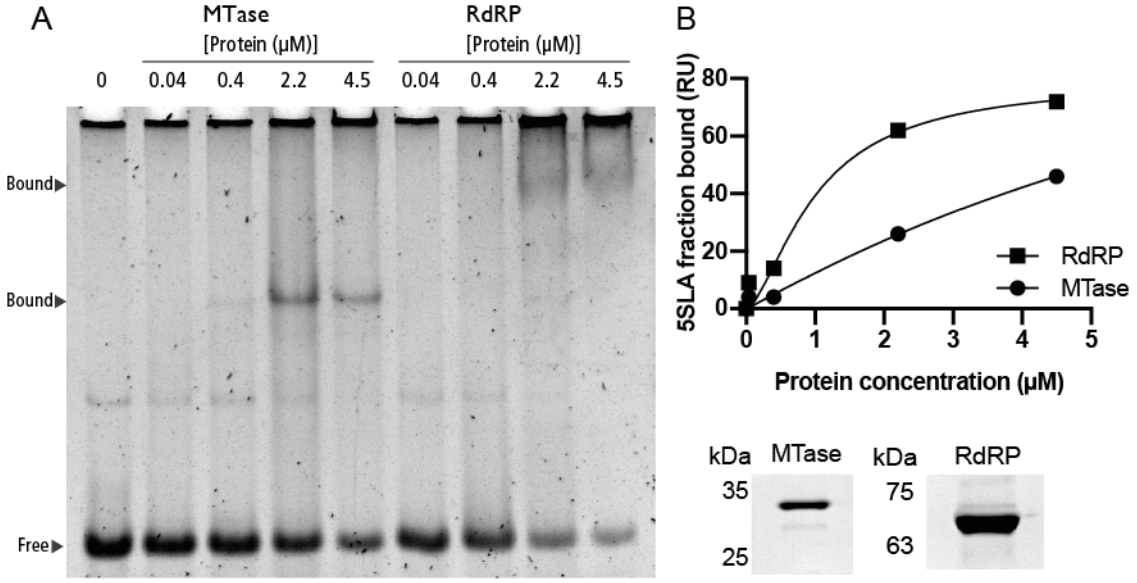


Figure 35: (A) Electrophoretic mobility shift assay (EMSA), showing the 5SLA binding to increasing concentrations of the individual NS5 domains. Protein concentrations are indicated above each lane. **(B)** Quantification of the 5SLA binding from (A). Data was obtained by integration of the bands corresponding to the free RNA and relativized to the first lane (F_{free}). Fraction corresponding to the bound 5SLA was calculated as $F_{bound}=1-F_{free}$ and ploted against the concentration of each domain. The solid lines correspond to the nonlinear regression of the 5SLA bound fraction to Hill equation. Obtained K_d and Hill slope (h) values were obtained from this equation and affinities were calculated as $K=1/K_d$. Binding affinities for the RdRP was K=0.99 μM (K_d=1.01 μM^{-1}, h=1.57, R^2=0.98) and the K=0,10 μM (K_d=9.87 μM^{-1}, h=1.05, R^2=0.99) in the case of the MTase.

Based on the structural information of the NS5-5SLA cryo-EM structure and the HDX data, we identified the NS5 residues involved in 5SLA binding in both A and B interfaces located at the MTase and RdRP domains respectively.

Within the interface A, we selected residues K226 and K231 within α-8 and its C-terminal loop, respectively. To confirm the role of these residues in 5SLA binding, we generated the NS5 mutant K226E, K231E that would disrupt the MTase-5SLA contacts.

Within the interface B, we selected residues K843 and R844 within the α-31 and R856 in α-32. In addition, previous studies described that residue Y840 was involved in binding the 3'SL RNA in DENV (Hodge *et al.* 2016). As ZIKV 5SLA and 3SL share the same top loop sequences (Figure 1 and 2), we decided to evaluate the relevance of this residue in 5SLA binding.

To confirm the role of all of these residues in the 5SLA binding, we generated several combinations of NS5 mutants, including the residue substitutions: K226E, K231E, Y840A, K843E, R844D and R856D to disrupt the interactions within these interfaces. To assay the stability of the NS5 5SLA binding mutants, the folding or all used mutants was assayed using DSF (DSF curves corresponding to the NS5 5SLA binding mutants can be consulted in Annex as Figure annex 2).

The NS5-5SLA binding capacity of all mutants was analyzed by EMSA and the 5SLA relative binding of each mutant was plotted (Figure 36). NS5 affinities were calculated for all the 5SLA binding mutants and the wild type NS5 (K=0.76 μM) (Figure 36). NS5-5SLA binding was impaired when the NS5 variant K226E and K231E in the MTase interface was assayed (K=0.51 μM). Similarly, all assayed NS5 variants to impair the 5SLA interactions at the RdRP interface resulted in a diminished total binding and affinities (Y840A-R856D: K=0.54 μM; Y840A-K843E-R844D-R856D: K=0.56 μM) with affinity reductions up to ~50% in the case of the residue substitutions K843E and R844D (K=0.41 μM) revealing an important role of these residues in the interaction with 5SLA (Figure 36). However, the NS5 that carried all mutations bound 5SLA with wild-type similar affinity (K=0.74 μM), although we could not completely abolish the interactions

using this mutant, it showed a reduced 5SLA binding at maximal NS5 concentration (Figure 36).

Hill coefficient (h) is a dimensionless parameter that indicates the “interaction” coefficient, reflecting the extent of cooperativity among multiple ligand binding sites and values greater than 1 is an indicative of positive cooperativity. Hill coefficient values were obtained for all mutants (Figure 36).

Wild type NS5 showed positive cooperativity in the 5SLA binding (h=2.60), and similar h values were obtained when the NS5 mutants on the RdRP 5SLA binding interface were assayed (K843E/R844D: h=2.23; Y840A/R856D: h=2.94 and Y840A/K843E/R844D/R856D: h=2.89). Despite NS5 mutant K843E/R844D registered the highest drop in the Hill coefficient of all RdRP mutants assayed, this effect was not observed in the other variant that carried the same mutated residues (Y840A/K843E/R844D/R856D). However, analysis of NS5 mutants on the MTase 5SLA binding interface showed decreased h values (K226E/K231E: h=1.82 and K226E/K231E/Y840A/K843E/R844D/R856D: h=2.18) (Figure 36).

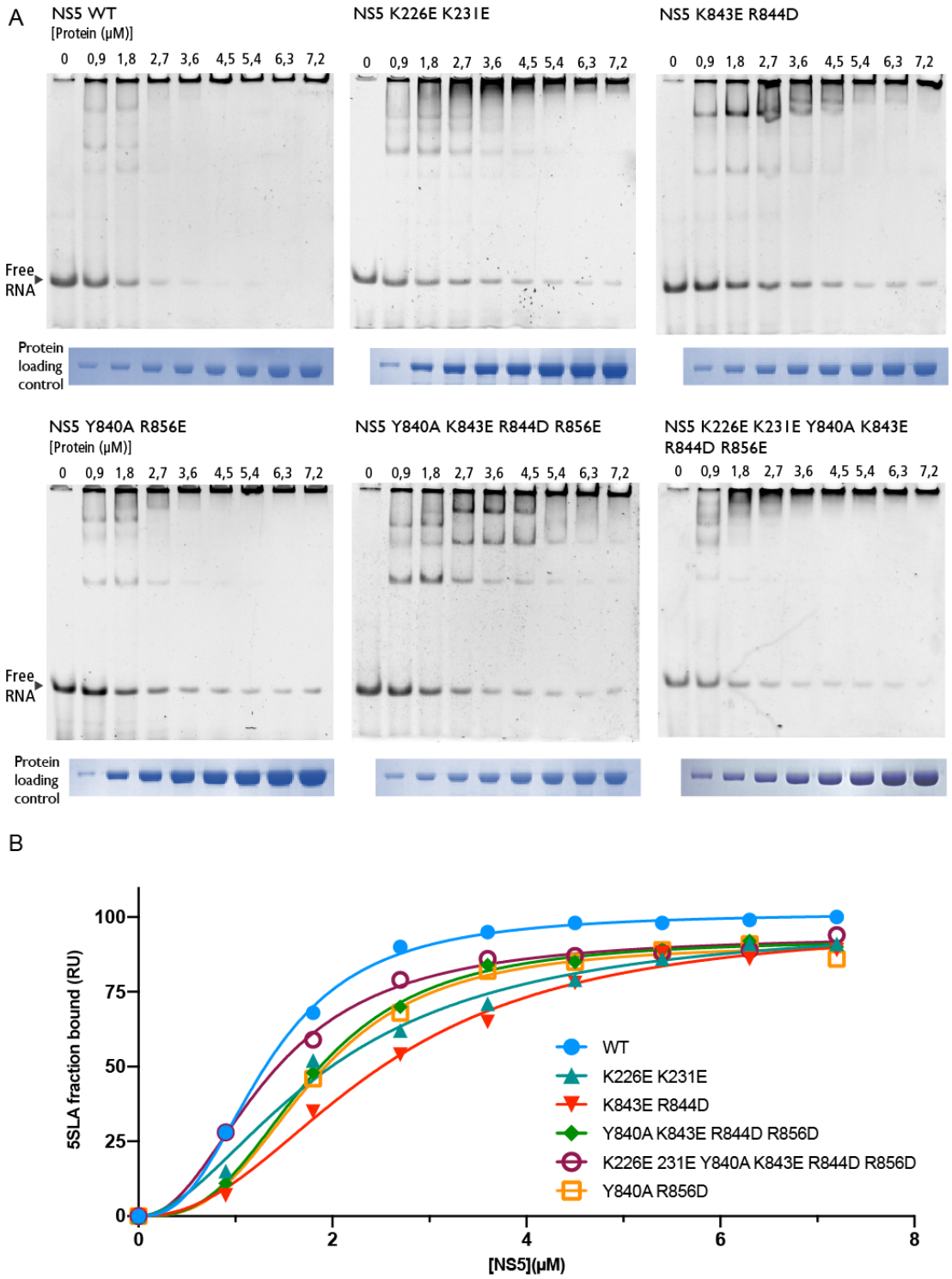


Figure 36: (A) Electrophoretic mobility shift assay (EMSA) results showing the 5SLA binding to different NS5 mutants on the interaction surfaces with 5SLA. NS5 protein concentrations are indicated above each lane. NS5 protein loading control for each EMSA is showed below the gels. **(B)** Quantification of the 5SLA binding to NS5. Data was obtained by integration of the bands corresponding to the free RNA in **(A)** and relativized to the first lane (F_{free}). Fraction corresponding to the bound 5SLA was calculated as $F_{bound}=1 - F_{free}$ and ploted against the concentration of each NS5 mutant. The solid lines correspond to the nonlinear regression of the 5SLA bound fraction to Hill equation. Obtained K_d and hill slope (h) values were obtained from this equation and affinities were calculated as $K=1/ K_d$. The wild type NS5 bound 5SLA with K=0.76 μM (K_d=1.32 μM^{-1}, h=2.06, R^2=0.99). NS5 mutant K843E-R844D exhibited the lowest 5SLA affinity K=0.41 μM

(K_d=2.47 μM^{-1}, h=2.23, R^2=0.99). NS5 mutants K226E-K231E, Y840A-R856D and Y840A-K843E-R844D-R856D bound 5SLA with similar affinities K=0.51 (K_d=1.97 μM^{-1}, h=1.82, R^2=0.99), K=0.54 (K_d=1.85 μM^{-1}, h=2.94, R^2=0.98) and K=0.56 (K_d=1.73 μM^{-1}, h=2.89, R^2=0.99) respectively. The NS5 variant that carried all mutations bound 5SLA with wild-type similar affinity K=0.74 μM (K_d=1.35 μM^{-1}, h=2.18, R^2=0.99) but showed a reduced 5SLA binding at maximal NS5 concentration.

The observed drop in the Hill coefficient when the mutations on K226E and K231E were assayed might indicate a cooperative binding between both MTase and RdRP domain. Binding affinities and Hill coefficient data also suggest that more complex interactions take place in the RdRP as the accumulative mutations did not altered the binding constants as expected, and other 5SLA binding residues could be leading the interaction.

3.2.9 Effect of 5SLA on NS5 polymerase activity

To assess whether the conformational changes observed in the cryo-EM structure affected the NS5 RdRP polymerase activity upon 5SLA binding, we employed a previously described fluorescence-based *in vitro* polymerization assay (Sáez-Álvarez *et al.* 2019) to measure *de novo* polymerization.

The *de novo* polymerization activity was measured using a Poly(U) (~1000 nt) as ssRNA template that would be first primed using the NS5 primer loop and then extended, yielding a dsRNA product. As result of these two activities, we would expect that NS5 was able to efficiently synthesize dsRNA from a ssRNA template. However, when NS5 was incubated in presence of 5SLA, the *de novo* polymerization activity was partially inhibited in a dose-dependent manner.

Analysis of the curves during the first minutes of polymerization where the *de novo* initiation take place showed subtle differences; curves that corresponded to higher 5SLA concentrations appeared shifted to the right. Moreover, calculated times for achieve the half of the maximal product were higher in all 5SLA curves. This suggest a retarded *de novo* polymerization activity during the first stages of the polymerization assay (Figure 37 A).

Polymerization reactions contained NTP, NS5 and 5SLA that was previously incubated 45 min before adding the PolyU RNA template as otherwise no differences in the activity were observed in 5SLA samples, probably because the generation of an intermediate NS5-RNA template that was masking the 5SLA effect.

These *de-novo* polymerization results could indicate that the ZIKV NS5 *de novo* priming activity is partially inhibited as consequence of 5SLA binding that would trigger the extrusion of the primer loop outwards the active site, impairing the formation of the replication-initiation complex (Potisopon *et al.* 2014) and in contrast, facilitating the elongation of a preformed template-primer duplex (Figure 33 C).

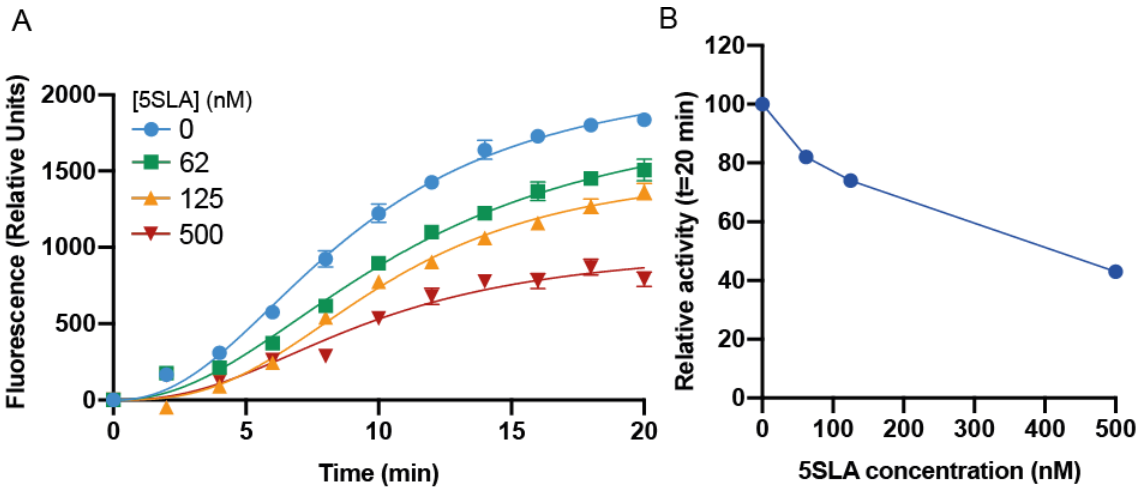


Figure 37: (A) Reaction progress plot of NS5 *de-novo* polymerization assay using a homopolymeric PolyU ssRNA template in presence of different 5SLA concentrations. Polymerization activity was recorded as the relative increase in fluorescence over 20 min. Solid lines represent non-linear regression with sigmoidal adjustment of the the data. Velocities (V) were calculated as the increment of fluorescence after 20 min and are expressed in relative fluorescence units per min (RFU/min). $T_{1/2}$ refers to the time needed to reach the half of the maximum product at t=20 min. V and $T_{1/2}$ for 0 nM 5SLA were 91.9 RFU/min and 9.1 min, for 62 nM 5SLA were 75.4 RFU/min and 10.1 min, for 125 nM were 68.3 RFU/min and 10.3 min, for 500 nM 5SLA were 40.0 RFU/min and 9.6 min. **(B)** Representation of the relative decreasing of V as function of the 5SLA concentration.

3.2.10 5SLA RNA crystallization

As our attempts to crystallize the NS5-5SLA complex failed, we wanted to gain insight from the 5SLA RNA. We performed several crystallization screening experiments using three different 5SLA constructs: 5SLA long, 5SLA short and 5SLA motif (Figure 38) (Table 2). RNA concentration was set to 8 mg/mL and the commercial crystallization screening Helix was placed at 4 ºC. After several days, the construct 5SLA long and 5SLA short crystallized in five crystal conditions exhibiting different crystal form (Figure 38), while no crystal appeared in the drops that contained the construct 5SLA Motif.

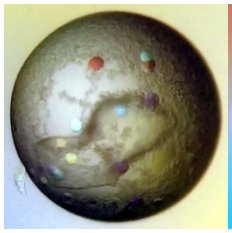
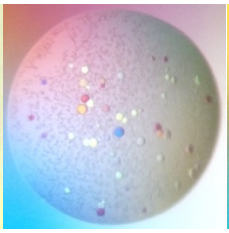
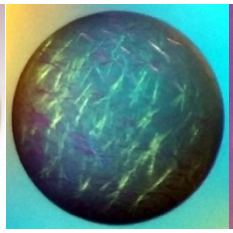
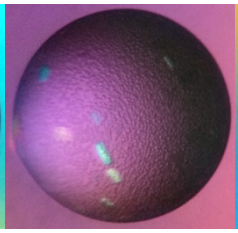
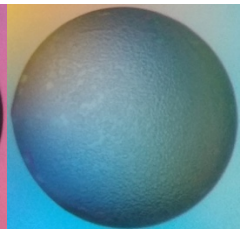
5SLA constructs crystallized				
5SLA Long	5SLA Long	5SLA Long and Short	5SLA Long	5SLA Long
				
25% PEG 400 0.1M Sodium Chloride 5 mM Strontium Chloride haxahydrate 50 mM MES pH 6.25	30% MPD 0.3 Sodium Chloride 50 mM HEPES pH 6.5	17% PEG 4000 0.1 M Lithium Chloride 10 mM Manganese (II) chloride tetrahydrate 50 mM MES pH 6.5	17% PEG 4000 0.1 M Sodium Chloride 50 mM Lithium Chloride 10 mM Magnesium Chloride hexanydrate 50 mM MES pH 6.5	17% PEG 4000 0.1M Sodium Chloride 5 mM Strontium Chloride Hexanydrate 50 mM MES pH 6.5

Figure 38: ZIKV 5SLA RNA crystal conditions at 8 mg/ml. RNA for crystallization were produced and purified in-house as described in M&M. For crystallization experiments, RNA constructions 5SLA long, 5SLA short and 5SLA motif (Table 2) were concentrated to 8 mg/ml but only 5SLA long and 5SLA short produced crystalls.

Cristal conditions were optimized separately and by making crossed combination in order to explore possible better crystal conditions. Finally, two optimized conditions yielded hexagonal-bipyramid crystals with good size and morphology in the crystallization conditions 25-30% PEG 4000, 0.1-0.2 M NaCl, in presence of 10 mM $MgCl_2$ or 5 mM Strontium Chloride Hexahydrate. X-Ray diffraction experiments of best crystals exhibited resolution limits at 10 Å (Figure 39) and several datasets were collected. Diffraction data was indexed with XDS as P1 with unit cell dimensions of a = 162 Å, b = 165 Å, c = 318 Å, however, after checking automatic spot detection, not all spots were correctly detected in the long axis (c=318), we also tested several indexing program as iMosfilm or Xia2Dials but obtained similar results. Due to the cell characteristics, the absence of models and the low resolution diffraction, we did not continue working on these crystals.

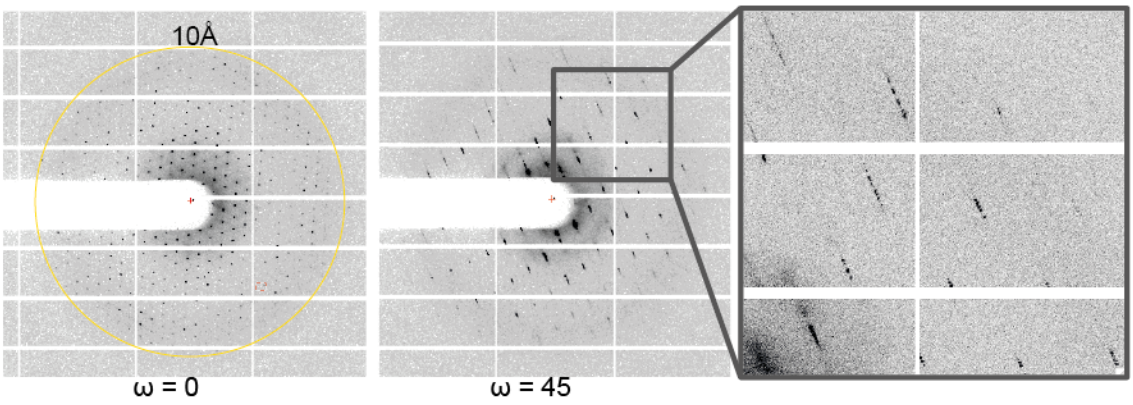


Figure 39: Diffraction pattern of the ZIKV 5SLA RNA at two different angles. Crystals showed diffraction spots at a resolution limit of 10 Å. Experiments were done using a diffraction wavelength of 1.03 Å.

4. Discussion

4.1 Structural, Biophysical and Biochemical characterization of ZIKV NS5

Genome replication in flavivirus takes places within replication factories (RF). It consists in a complex network of rearranged membranes derived from the ER were the non-structural proteins, hosts factors and viral genome meet in a concerted form allowing viral replication (Klema *et al.* 2016, Paul *et al.* 2013). Electron tomography studies of different flaviviruses RF reveal the presence of induced invaginations in the ER to form clusters of vesicles packets (VPs) where viral replication takes place (Welsch *et al.* 2009).

The VPs contain double-stranded RNA (dsRNA), which is an intermediate product of viral genomic RNA replication (Den Boon *et al.* 2010). Therefore, it seems that RNA synthesis occurs inside these vesicles. The presence of small pores connecting the vesicle membranes may be used for exporting the newly synthesized genomic RNA to the site of protein synthesis or to the site of particle formation (Arakawa and Morita 2019). Many aspects such as protein composition, distribution or stoichiometry within RF during flavivirus replication remain still unclear. However, several studies described the intracellular accumulation of viral polymerases during viral infection in both positive (+) and negative (-) ssRNA viruses such as Poliovirus (Hobson *et al.* 2001), Human Norovirus (Högbom *et al.* 2009), Sendai virus (Smallwood *et al.* 2002) or Rift Valley Fever virus (Zamoto-Niikura *et al. 2009)*, uncovering the biological relevance of these structures for viral replication. Furthermore, the impact of the oligomeric state on the *in-vitro* enzyme activity has been studied in several virus families, including the *Flaviviridae* family, revealing different regulatory mechanisms and functions associated to this process (Choi *et al.* 2004, Chinnaswami *et al.* 2010, Clemente-Casares *et al.* 2011, Ferrero *et al.* 2015).

The crystal structures of ZIKV NS5 reported in this thesis work reveal a dimer based helical arrangement of NS5 molecules extending along the crystal c-axes (Figure 13). Head to tail NS5 dimers (type I interactions) are formed by molecules AB, CD and EF in the $P2_12_12_1$ crystals and by A'B', C'D' and E'F' in the $P6_5$ crystals (Figure 7). The presence of type I NS5 dimer has been confirmed by

several biophysical techniques such as SEC-MALS, AUC and SAXS (Figures 14 and Figure 15) and the contacting interface has also been confirmed by crosslinking experiments coupled to MS (Figure 16 and 17). Crosslinking experiments further validated that MTase domain was the responsible of driving the NS5 dimerization and mutagenic studies revealed that residues Y25, K28 and K29 are essential to form this interaction (Figure 17). Dimer-dimer contacts (type II interactions) are stablished between molecules BC and DE, and B'C' and D'A'', in the $P2_12_12_1$ and $P6_5$ crystals respectively (Figure 7). The combination of type I and type II contacts results in the formation of the large helical structures in the packing of both crystal forms (Figure 13) and would be the responsible for the presence of the fibril-like structures observed by both negative-staining EM and AFM (Figures 18 and 19). The lack of a regular pattern of the fibril-like structures observed by negative staining EM and AFM could be due to the intrinsic interdomain flexibility of the NS5 molecule, as seen in our SAXS experiments (Figure 15) and also described for ZIKV and DENV (Saw *et al.* 2017, Bussetta *et al.* 2015). Other recently published structures for DENV revealed alternative orientations of the MTase domain (Sahili *et al.* 2019, Wu *et al.* 2020), confirming that the linker between MTase and RdRP facilitates the movement of one domain with respect to the other.

Interestingly, the fiber-like structures were not observed when the NS5-Y25A/K28S/K29A variant was imaged (Figure 18 and 19). Therefore, the combined structural and biophysical techniques applied to the isolated NS5 shows that the protein is able to form stable dimers in solution and these dimers may act as building blocks for the formation of higher-order oligomers that could play a role in the regulation of the multiple enzyme functions in the context of the replication complex.

NS5 oligomeric arrangements may also mimic the association of the enzyme with other viral and/host factors in the RFs. For instance, the helicase NS3 protein that unwinds the dsRNA and removes the ɣ-phosphate at the 5'-RNA prior to cap addition. It is expected that individual activities of both NS5 and NS3 would be modulated by mutual interactions. Their respective binding sites map to the C-terminal domain of NS3 (residues 566-595) (Tay *et al.* 2015) and a region on the RdRP fingertips, containing the conserved charged residue K331 that also

appears to play a critical role in nuclear import (Johansson *et al.* 2001, Zou *et al.* 2011). In the ZIKV NS5 dimeric form, the NS3 binding site in the RdRP domain is also in close proximity of the neighbor MTase and might facilitate the combined activities (Figure 7).

In the ZIKV NS5 structure, the active sites of MTase and RdRP lay in opposite faces and do not interact with each other. A similar conformation was described for JEV and DENV NS5 protein (Figure 11), and large conformational changes were predicted to occur in order to facilitate the MTase to occupy a position, near of the recently synthesized dsRNA allowing the 5'-RNA cap transfer (Selisko *et al.* 2014). The strictly conserved GTR motif at the N-terminus of the MTase-RdRP linker was proposed to act as a hinge, allowing the required interdomain movements (Klema *et al.*2015, Selisko *et al.* 2014).

An alternative hypothesis implying the coordination of both MTase and RdRP activities was proposed for NS5 dimers observed in DENV (Klema *et al.* 2016). In DENV model, the RdRP dsRNA exit channel in one monomer and the MTase active site of its partner were oriented in the same direction, suggesting a coordination mechanism between MTase and RdRP domains without further conformational changes (Klema *et al.* 2016). Similar to that reported for DENV, in ZIKV NS5 type I dimers (Figure 7 A) the dsRNA exit site of one protomer faces the MTase active site of the adjacent protomer in the dimer (Figure 40). This neighbor MTase is in closer proximity to the NS3 binding site in the partner RdRP (14 Å) than in its own RdRP domain (30 Å), supporting the idea that the different combined activities (dsRNA unwinding, removal of the 5'-RNA ɣ-phosphate and cap addition) could be carried out in trans (Figure 40).

Our *in-vitro* elongation assays revealed that the NS5-Y25A/K28S/K29A variant, displaying impaired dimerization capacity, was more efficient in primer extension than the wild type enzyme, this effect was specially seen at high concentrations at which WT NS5 dimerization is favored (Figure 21). This result suggests a possible regulatory mechanism by which NS5 dimerization negatively regulates RNA synthesis and chain elongation during viral infection.

Furthermore, besides of being essential for dimerization, NS5 residues K28 and K29 are also important for GTP binding (Egloff *et al.* 2002). Our competition assays showed that NS5 dimerization was hindered at high GTP concentrations (Figure 19) suggesting a regulation in addition, of the capping process.

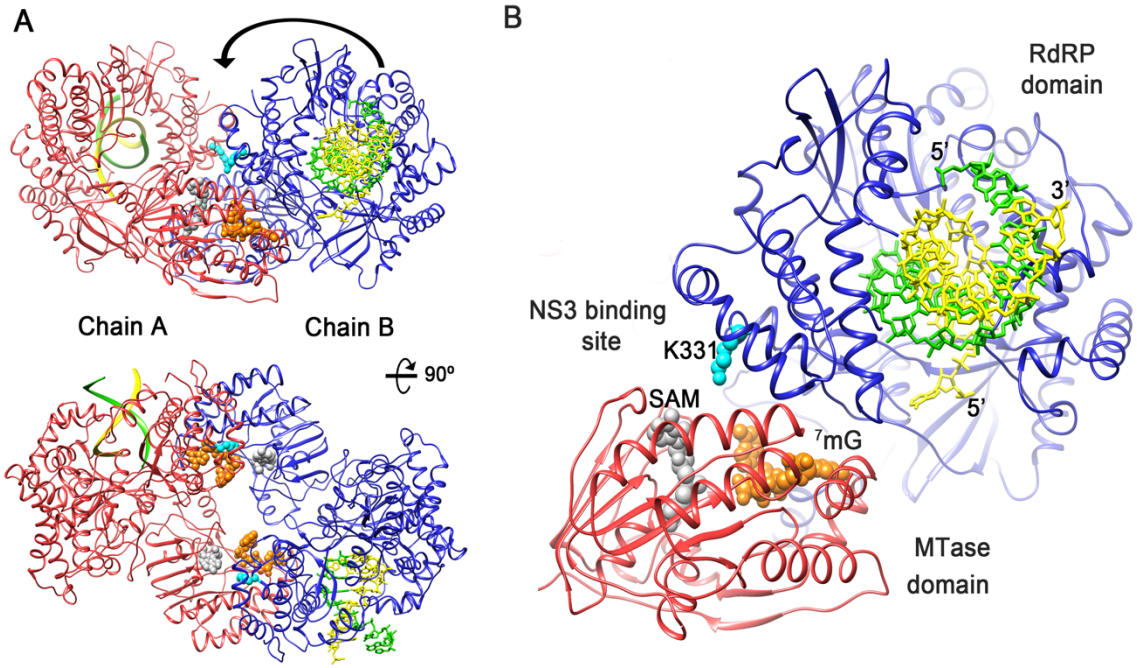


Figure 40: Cooperative model of RNA syntesis. **(A)** Cartoon representation of two views of the NS5 dimer with monomer A shown in red and monomer B in blue. A RNA template/primer is modeled in the polymerase central cavity of the two RdRPs, based in the structure of a PV replication-elongation complex (PDB: 3OL6), with the template strand shown in yellow and primer strand in green. The SAM cofactor (white spheres) and the ^{7}mG molecule (orange spheres) have been modeled in the MTase active site by superimposition of the DENV MTase structure (PDB: 3EVE). The RdRP domain residue K331, essential for NS3 binding is shown as spheres in cyan. **(B)** Close up view highlighting the direct connection between the dsRNA exit of one monomer (blue with RNA in yellow and green), the NS3 binding site (cyan spheres) and the entrance to the MTase active site of its partner (red) in the dimer.

4.2 The multimeric arrangement of ZikV-NS5 interferes with the molecular machinery and behavior of Neural Progenitor Cells

By far, NS5 has only been shown to counteract the host defense mechanism by inhibiting interferon pathway (Grant *et at.* 2016). *In-vivo* experiments, performed in collaboration with the group of Drs. Elisa Martí and Murielle Saade (Department of Developmental Cell, IBMB), revealed that NS5 has a toxic effect on NPCs by inducing premature NPC differentiation (Figure 22). We also demonstrated that the altered mode of division of ZIKV infected NPCs was strongly dependent of the NS5 quaternary structure. As shown in Figure 20, the presence of ZIKV NS5 on the neural tube of chick embryos induces premature neuron differentiation,

resulting in the exhaustion of NPCs pool. We have also demonstrated that while the enzymatic activities of ZIKV NS5 appear to be dispensable, the Y25, K28 and K29 residues in the protein that are involved in NS5-oligomerization, are essential for promoting premature neurogenesis (Saade *et al.* 2019). This data suggests that NS5 is directly involved in the developmental malformations attributed to congenital ZIKV syndrome.

4.3 The NS5-5SLA complex

Flavivirus genomes are dynamic structures that acquire different conformations along the viral cycle. Local structures and long-range RNA-RNA interactions regulates viral replication and interact with both viral and host proteins for proper viral replication. Several studies described the interaction of 5SLA with NS5 by biochemical, biophysical and predictive models (Bujalowski *et al.* 2017, Bujalowski *et al.* 2020, Hodge *et al.* 2016, Filomatori *et al.* 2006). However, there are no three dimensional structures of either the 5SLA RNA nor the 5SLA-NS5 complex determined so far.

In this work we evaluated the binding of this NS5-5SLA complex using analytical SEC and EMSA assays (Figure 24). We also optimized the buffer conditions to carry out cryo-EM experiments, using DLS, EMSA and DSF (Figure 25). We have obtained the cryo-EM structure of ZIKV NS5 in complex with the 5SLA promoter at 9.4 Å of resolution (Figures 28, 29 and 30) (Table 3).

Our Cryo-EM maps showed a monomeric ZIKV NS5 protein bound to the 5SLA RNA structure by both MTase and RdRP domains. The two RNA contacting surfaces in NS5 involve a large number of electropositive residues (Figure 30). Cryo-EM data was further validated by HDX-MS experiments, showing that residues 222-240, comprising β-8 strand and α-8 helix in the MTase domain, and residues 828-848 in α9, and 854-872 in α10, located within the RdRP thumb subdomain, are in close contact with the 5SLA (Figure 34). Previous site directed mutagenesis experiments experiments performed in DENV identified similar interactions between the RdRP domain of NS5 and the 3SL RNA top loop structure (Hodge *et al.* 2016). Substitutions in DENV NS5 residues R770, R773,

Y838 and K841 resulted in diminished 3SL binding and impaired viral replication (Hodge *et al.* 2016).

Site directed mutagenesis coupled to EMSA assays allowed us to validate the role of residues K226 and K231 in the MTase, and residues Y840, K843, R844 and R856 in the RdRP, in 5SLA binding (Figure 36). Interestingly, in ZIKV, both 5SLA and 3SL structures share the same top loop sequence (5' ACAG 3') and the 3SL RdRP interaction interface described for DENV (Hodge *et al.* 2016), matches the same surface as in our ZIKV NS5 5SLA cryo-EM structure, prompting to hypothesize that both 5SLA and 3SL promoters share the same binding pocket in the RdRP thumb subdomain.

Very recently, it has been reported the characterization of the ZIKV 5SLA binding affinities to the full-length NS5 and to their individual RdRp and MTase domains. In addition, site-directed mutagenesis with specific positively charged residues on NS5 both domains (Bujalowski *et al.* 2020). Data reported in this work coincides with our mutational data (Figure 36).

One of the most interesting findings of the NS5-5SLA cryo-EM structure is that despite the 5SLA molecule does not interact directly with the RdRP active site, its binding induces large rearrangements in the RdRP domain, resulting in a widening of the central channel were the catalytic site is located (Figure 31). In addition to the rotation of the thumb sub-domain with respect to the fingers, the EM maps also showed an extra density at the thumb subdomain domain that could be interpreted by the displacement of the priming loop structure from its position in the unbound structure, occluding the active site to an extruded conformation that allows the opening of the central channel (Figures 31, 32 and 33).

In the *de novo* synthesis, one initiation nucleotide provides the 3′-hydroxyl, serving as a primer for the addition of the next nucleotide to form the first phosphodiester bond of the product strand. Structural comparisons showed that RdRPs of viruses using *de novo* initiation (i.e. within the *Flaviviridae* family) share distinctive characteristics, ensuring competent replication initiation, including the

priming loops that fill most of the active site cavity and provide a platform for the initiating nucleotides. These platforms also serve as a barrier that prevent chain elongation (reviewed in Lescar and Canard 2009, Ferrer-Orta *et al.* 2006; Ferrero *et al*. 2018). Therefore, important conformational changes, moving the priming loop away from the active site, would be required to allow the transition from the initiation to the elongation states of RNA synthesis (Mosley *et al*. 2012, Appleby *et al*. 2015).

The biochemical experiments performed in this thesis indicate that the *de novo* initiation of RNA synthesis is considerably reduced in presence of increasing concentrations of 5SLA (Figure 37). These results are not surprising in light of the structural data, showing the movement of the priming loop away the active site, impairing the formation of the replication-initiation complex and, in contrast, facilitating the elongation of a preformed RNA template-primer duplex (Figures 31 and 37).

Similar rearrangements were observed when the crystal structures of HCV NS5B unbound and dsRNA bound forms were compared. Mosley and co-workers removed the priming loop residues from a recombinant HCV NS5B and observed a strong reorganization of thumb subdomain that appears rotated with respect to the fingers, opening the active site cavity (Mosley *et al.* 2012). Later works also described higher temperature factors within the priming loop, suggesting this structure started to become disordered during the primed-initiation (Appleby *et al.* 2015). Priming loop deletion in this work enabled obtaining the structure of the NS5B-RNA complex in elongation state, showing the dsRNA occupying the polymerase central cavity (Appleby *et al.* 2015).

Similar conformational changes in the thumb subdomain and extended primer loop upon 5'UTR binding were also observed in RdRPs of negative-sense viruses such as Influenza and Lacrosse viruses (Kouba *et al.* 2019, Wandzik *et al*. 2020, Arragain *et al.* 2020) suggesting conserved regulatory mechanisms for the transition from initiation to elongation processes in the *de novo* replication of viral RNA genomes.

The obtained cryo-EM structure confirms the previously described NS5-5SLA interactions for DENV and is compatible with the proposed model for flavivirus

genome transcription (Filomatori *et al.* 2006). In this model, once the ZIKV genome is circularized, both 5' and 3' UTR end regions would end up very close one to each other, and as a consequence of circularization, the genome 3' end would adopt an unstructured conformation leaving an available 3' end as template to initiate viral replication (Figure 1).
In the cellular environment, NS5 RdRP would synthetize the first two incoming nucleotides aimed by the priming loop (Figure 41). NS5 binding to the circularized genome conformation also allows binding of both the unstructured 3' end at the template channel and the 5' 5SLA promoter at the thumb subdomain in a proper conformation for RNA synthesis.

So far, this is the first flavivirus structure describing the interaction between NS5 and the 5SLA promoter. As a result of the moderated resolution we obtained for this complex, we cannot describe the atomic structure of this complex with confidence. However, we can clearly confirm that the RdRP domain undergoes through conformational changes that end up in central channel widening and primer loop extrusion. Our biochemical assays have confirmed the observed intermolecular interactions that drive this complex and shed new light on the mechanisms of regulation of RNA synthesis in ZIKV NS5. The NS5-RNA interactions observed in this complex, in particular those involving the thumb subdomain, can be targeted for the design of new antivirals perturbing viral replication.

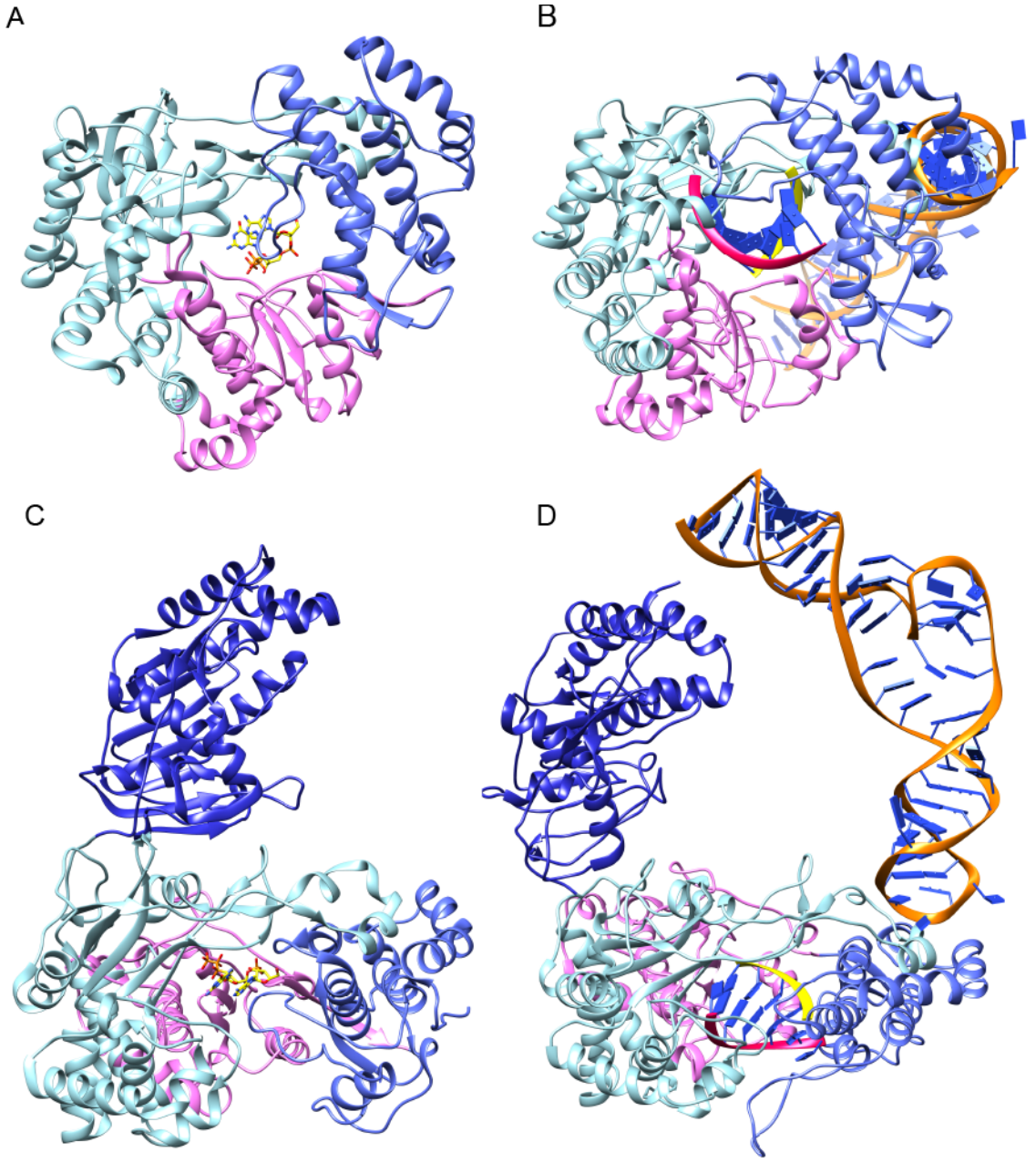


Figure 41: Proposed model for ZIKV NS5 *de novo* genome initiation and elongation. **(A)** Aimed by the priming loop, NS5 would synthetize the first di-nucleotide (in yellow) that would be used for priming the ssRNA 3' genome end. **(B)** Upon binding the 5SLA (in orange) of the circularized genome and the 3' genome end, NS5 would adopt a pre-elongation state where the template strand (colored in red) would be primed and the first nucleotides are added. In this state, NS5 would displace the priming loop in order to drive the dsRNA (colored in yellow and red) out. Panels **(C)** and **(D)** correspond to top views of (A) and (B). Model from (A) and (C) corresponds to the NS5 crystal structure (6I7PB) while models in (B) and (D) correspond to the NS5-5SLA cryo-EM model. The di-nucleotide primer was modelled using chimera. The model corresponding to the dsRNA was taken from the HCV elongation complex PDB: 4WTA.

5. Conclusions

1. The crystal structure of the full-length ZIKV NS5 shows that the protein adopts an extended fold in which the MTase domain (residues 1–262) is placed above the fingers sub-domain, behind the RdRP (residues 273–888), opposite to the double stranded RNA exit channel and close to the template binding channel. A conformation closely related to that found in NS5 of JEV, a closely related member of the Flaviviridae family.

2. Analysis of the crystal packing interactions in two different space groups (orthorhombic $P2_12_12_1$ and hexagonal $P6_5$) allowed the characterization of the supramolecular arrangement of the NS5 protein, highlighting the assembly and interfaces between NS5 monomers within a dimeric structure, as well as the dimer-dimer interactions to form higher order fibril-like structures.

3. Combined structural, biophysical and functional data confirm the ability of the ZIKV enzyme to form stable dimers in solution as the building blocks for the formation of higher-order oligomers that would participate in the fine-tuned regulation of the multiple enzyme functions in the replication complex.

4. NS5 dimerization is leaded by the MTase domain with residues Y25 K28 and K29 playing a crucial role in dimer formation.

5. NS5 dimerization is disrupted at high GTP concentrations, providing an explanation of how dimerization might regulate the capping process.

6. The NS5 dimerization down-regulates the RdRP polymerase activity of ZIKV NS5, as the NS5 variant Y25A/K28E/K29E that showed impaired dimerization also exhibited an increased elongation activity, when compared with the wild-type NS5.

7. ZIKV-NS5 produces an imbalance in the mode of division of NPC, triggering premature neurogenesis and in consequence a diminished number of neurons, as the main cause of microcephaly. While the enzymatic activities of NS5 appear to be dispensable, the Y25, K28 and K29 residues, involved in NS5-oligomerization, are essential for premature neurogenesis.

8. ZIKV NS5 binds the 5SLA structure of the 5'UTR through the MTase and RdRP domains and residues K226, K231, Y840, K843, R844 and R856 play a role in this this binding.

9. The NS5 binding to 5SLA induces a conformational change on NS5 that includes movements in both MTase and RdRP domains. The largest changes occur in the RdRP where a widening of the central channel and the displacement of the primer loop outwards take place.

10. The NS5 binding to 5SLA triggers the structural changes required for the transition from a conformation able to initiate *de novo* RNA synthesis where the priming loop serves as platform for the positioning of the initiating nucleotides, to a conformation compatible with chain elongation where the central cavity is opened to facilitate the exit of the recently synthesized dsRNA.

6. Materials and methods

6.1 NS5 cloning expression and purification

ZIKV NS5 protein coding sequence (Suriname strain isolate Z1106033, GeneBank accession number KU312312.1) containing a GSSS linker and an additional 6xHis C-terminal tag at C-terminal including flanking sequences which provide NdeI and EcoRI restriction sites at the 5' and 3' ends respectively was designed. The construct was codon optimized for bacterial expression, commercially synthesized and cloned into a pMK backbone plasmid. To recover NS5 coding sequence, the plasmid was digested with NdeI and EcoRI restriction enzymes and ligated with T4 ligase into a previously NdeI/EcoRI pRSET linearized plasmid (ThermoFisher). The resulted plasmid pRSET ZIKV NS5-His, carries at 5' a T7 promoter sequence and encodes the full-length ZIKV NS5 protein fused to a C-terminal tag (NS5-His). This vector was used as template for further constructs and punctual mutations.

An aliquot of fresh BL21(DE3) *E.coli* competent cells were transformed by a heat-shock for 1 min at 42 ºC. A 75 mL LB culture was set up by picking three single colonies and grown on agitation O/N at 37 ºC and used as starter to inoculate 650 mL LB flasks that were grown at the same conditions until a OD of 0.6-0.9 was reached. Cell cultures were then induced with the addition of 0.1 mM IPTG, and incubated at 200 rpm O/N at 18 ºC. *E.coli* cells were harvested and pellets were stored until use at -20 ºC.

Cell pellets of ZIKV NS5-His protein were lysed with homogenization buffer (50 mM Tris pH 7.0, 500 mM NaCl, 10% Glycerol, 1 mM DTT, 20 mM Imidazole) using a french press at 1.4 bar at 4 ºC, lysates were centrifuged at 18.000 rpm with a JA25.50 rotor for 45 minutes at 4 ºC and soluble fraction was filtered with a 0.22 µM syringe filters to remove cell debris. A 5 mL Nickel Inmobilized Metal Affinity Column (General Electrics) was equilibrated with homogenization buffer before apply protein sample, an extensive wash with the same buffer were performed to remove unbound sample and then, recombinant protein was eluted with a linear gradient of 500 mM Imidazole. Fractions that showed an increased absorbance at 280 nm were collected, pooled together and subjected to a dialysis against 50 mM Tris pH 7.0, 300 mM NaCl, 10 % Glycerol, 1 mM DTT O/N at 4

ºC. Dialyzed samples were then diluted to 100 mM NaCl with 50mM Tris pH 7.0, 10 % Glycerol 1 mM DTT and submitted to a cation exchange chromatography using a 5 mL HiTrap SP (General Electrics) column equilibrated with 50mM Tris pH 7.0, 100 mM NaCl, 10 % Glycerol, 1 mM DTT and eluted with a linear gradient of 1 M NaCl. Fractions that showed an increased absorbance at 280 nm were collected and analyzed by SDS-PAGE to evaluate the purity. Best fractions were pooled together and submitted to a size exclusion column (Superdex 200 10/200) in stabilization buffer (50 mM MES pH 6.0, 500 mM NaCl, 10% Glycerol, 5 mM DTT). Fraction that corresponded to the monomeric NS5 were pooled togheter and concentrated to 10 mg/ml using a 30 K centrifugal device (Millipore). Concentration was measured using a BioDrop quantification system and an extinction coheficient at 280 nm of 224710 M^{-1} cm^{-1}. Protein preparation was aliquoted, flash frozen in liquid nitrogen and stored at -80 ºC until use.

6.2 NS5 crystallization and X-Ray data collection

Initial crystallization assays of NS5-His were performed by the sitting-drop vapor diffusion method at both 277, 290 and 293 K in 96-well Greiner plates using a nanolitre-drop crystallization robot (Cartesian) and several crystallization screen at the Automated Crystallography Platform (PAC) located at IBMB-CSIC, Barcelona. Protein solution at 5 and 10 mg/mL in 50 mM MES pH 6.0, 500 mM NaCl, 10% Glycerol, 5 mM DTT was used to set up 96-well screen plates in which 100 nL of protein solution were mixed with 100 nL of precipitant solution. Plates were incubated at different temperatures and checked regularly. After 24h, first small bar-like crystals appeared on three crystal conditions: a) 0.2M Lithium Sulphate Anhydride, 0.1 CAPS pH 10.5, 1.2 M Sodium Dihydrogen Phosphate, 0.8 M Dipotassium Hydrogen Phosphate, b) 0.1 M Sodium Acetate pH 4.5, 0.8 M Sodium Dihydrogen Phosphate, 1.2 Dipotassium Hydrogen Phosphate and c) 1.8 M Sodium/Potassium Phosphate pH 6.9. After several rounds of optimization in 96-well plates, orthorhombic and hexagonal based prism crystals were obtained by mixing 300 nL of NS5-His protein solution with an equal volume of reservoir solution (0.8 – 1.15 M Sodium Dihydrogen Phosphate, 0.9 – 1.2 M Dipotassium Hydrogen Phosphate, 0.1 M Sodium Acetate pH 4.5 – 5.4). Best crystals were soaked in reservoir solution plus 20% (v/v) Glycerol before being harvested in cryo-loops and then cryo-cooled in liquid nitrogen.

Given that the first diffraction data obtained were at low resolution, crystal growth conditions and cryo-cooling were optimized. For this purpose, different salt concentrations in harvesting buffer (additional precipitant concentration, between 5 to 20%), capillary grown crystals as well as different types of cryoprotectant agents (glycerol, ethylene glycol, paraffin, mineral oil, silicon and sugars) were studied. Even crosslinking of crystals by vapor diffusion with 2% (v/v) of glutaraldehyde at different times were assayed without success. After optimization, the best diffracting crystals were harvested in a cryo buffer containing 5% higher concentration of phosphate salts than in the crystallization condition and 25% of glycerol (for orthorhombic crystals) or a mix of 4% sucrose, 2% glucose, 10% glycerol and 8% ethylene glycol in addition to the 5% of higher salt concentration (for hexagonal crystals).

Diffraction experiments of both obtained types of crystals were carried out on ID-23-1, ID-23-2 and ID-30A beamlines at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and at XALOC beamline at the ALBA Synchrotron (Cerdanyola del Vallès, Spain). All datasets were collected at 100 K. Diffraction data was processed using XDS. All datasets collected from orthorhombic crystals were pairwise-compared with SCALEIT (CCP4 suite), in a systematic manner, using a home-made script. Eleven partial datasets from different crystals were subsequently merged with XSCALE. Space group determination and data reduction were performed with POINTLESS/AIMLESS. This processing resulted in a complete data set at 5 A resolution that was used to solve the structure of NS5 by molecular replacement. A subsequent inspection of the data using the STARANISO software, currently available in the web server (http://staraniso.globalphasing.org), revealed that these crystals were highly anisotropic with diffraction limits of ellipsoid fitted to diffraction cut-off surface of: 5.85, 7.44, and 3.76 Å along the a*, b*, and c* axes, respectively. STARANISO analyses the anisotropy decay of the mean intensities (I), constrained by the crystal symmetry, modifies the TRUNCATE procedure and analyses the decay of the local average of I/σ(I) in different directions, providing the basis for an anisotropic resolution cut-off. The criteria used in determination of the resolution cut-off was: Rpim = 0.6, I/σI = 2.0 and CC1/2 = 0.3. When the anisotropy correction was applied to the unmerged data after cut-off produced a best resolution limit of 3.98 Å in direction 0.071 a* + 0.010 b* + 0.997 c* and a lowest

cut-off limit of 7.39 Å in direction 0.857 b*+0.514c*. The diffraction intensities from a single hexagonal crystal were scaled and merged using SCALA, resulting in a complete data set at 4.8 Å.

6.3 NS5 Structure determination

The $P2_12_12_1$ structure was solved by a combination of molecular replacement with PHASER, using two independent search models derived from the ZIKV MTase (PDB: 5KQR) and JEV RdRP (PDB 4K6M) domains, and manual fitting after density modification with SHELXE, starting from the coordinates of the partial solution yielded a map where the missing domains could be manually fitted. A global NCS was used during the initial refinement with Refmac5, owing to the presence of six copies of the NS5 monomer in the crystal asymmetric unit. Subsequent refinement rounds were performed by Phenix, using the 4 Å data corrected by STARANISO and the NCS was relaxed to a torsional NCS with groups automatically defined by the program. This refinement with Phenix was alternated with manual model rebuilding performed with Coot. The $P6_5$ structure was solved by molecular replacement with PHASER, using one entire ZIKV NS5 monomer from the orthorhombic structure as a search model. Refinement was done alternating cycles of refinement in Refmac5, applying torsional NCS and secondary structure restraints, with manual rebuilding using Coot. Superimpositions of structures were carried out using Coot. Calculation of contact surfaces were performed with PISA.

6.3 Small Angle X-Ray Scattering

NS5 samples were centrifuged at 10.000 g for 10 minutes at 4 ºC and diluted from a 10 mg/mL stock with the stabilization buffer (50 mM MES pH 6.0, 500 mM NaCl, 10% Glycerol, 5 mM DTT) to prepare 50 µL samples that were charged into the beamline autoloader. All SAXS measurements were done at the BioSAXS beamline (BM29, ESRF, Grenoble, France), An automatic pipeline already installed in the beamline was used to clean and load the vial with protein samples for X-Ray exposure, a total of 10 frames of 5s exposure at 10 ºC were performed with the same sample and best curves were used for software analysis. Scattering curves were calculated by subtracting the scattering buffer curves using ATSAS package (Franke *et al.* 2017). Data quality was assessed

for aggregation using Guinier plots, and Kratky plots were used to assess protein folding. The radius of gyration (Rg), experimental maximum particle diameter (Dmax) and molecular weight (MW) were calculated in Primus and the experimental molecular weight was estimated using the Porod Volume (Konarev *et al.* 2003). *Ab initio* models were generated using GASBOR (Svergun *et al.* 2001), for *ab*-initio model corresponding to scattering curve at 6 mg/ml, a P2 simmetry was imposed. To determine the conformational flexibility of the NS5 monomer, we used the ensemble optimization method (EOM). An ensemble of 10,000 random conformations was generated by keeping the RdRP domain constant and allowing the MTase domains (residues 1-268) to adopt any random conformation. EOM selected the optimized sub-ensemble from the random pool.

6.4 NS5 BS3 crosslinking and mass spectrometry

Crosslinking reactions were assembled at room temperature using 5 µg of purified NS5 protein in 50 mM HEPES pH 7.0, 150 mM NaCl in a final volume of 10 µL. As crosslinker agent a final concentration of 0.01 mM Bis-Sulfosuccinimidyl-Suberate (BS3) was added to the crosslinking reactions that were inmediately incubated for 45 min at 4 ºC and then stopped by adding 1M Tris pH 7.0 to a final concentration of 50 mM. In nuceotide-mediated dimerization inhibition reactions a final nucleotide concentration of 50 mM was used. To evaluate the presence of dimers, samples were mixed with 1X SDS loading buffer, heated for 5 minutes at 95 ºC and loaded into a 7.5% SDS-PAGE. Gels were run at a constant voltage of 100 V until ladder resolving and then stained with coomasie-blue. For crosslinking experiments coupled to mass spectrometry a band corresponding to NS5 dimer was carefully excised from the rest of the gel and trypsinized directly in polyacrylamide gel. The nano-LC-MS/MS set up was as follows. Digested peptides were diluted in 3% CH3CN, 1% Formic Acid. Samples were loaded to a C18 Trap column (Thermo Scientific) at a flow rate of 15µl/min using a Thermo Scientific Dionex Ultimate 3000 chromatographic system (Thermo Scientific). Peptides were separated using a C18 analytical column (Acclaim PepMap® RSLC 75 µm × 50 cm, nanoViper, C18, 2 µm, 100Å, Thermo Scientific). The column outlet was directly connected to an Advion TriVersa NanoMate (Advion) fitted on an Orbitrap Fusion LumosTM Tribrid (Thermo Scientific). The mass spectrometer was operated in a data-dependent

acquisition (DDA) mode. Survey MS scans were acquired in the orbitrap with the resolution (defined at 200 m/z) set to 120,000. The lock mass was user-defined at 445.12 m/z in each Orbitrap scan. The top speed (most intense) ions per scan were fragmented by HCD and detected in the orbitrap (resolution set to 30,000). The ion count target value was 400,000 and 50,000 for the survey scan and for the MS/MS scan respectively. Target ions already selected for MS/MS were dynamically excluded for 30s. Spray voltage in the NanoMate source was set to 1.70 kV. RF Lens were tuned to 30%. Minimal signal required to trigger MS to MS/MS switch was set to 50,000. The spectrometer was working in positive polarity mode and singly charge state precursors were rejected for fragmentation. At least two blank runs before of analysis were performed in order to ensure the absence of cross contamination from previous samples.

6.5 Atomic Force Microscopy

Measurements were performed at the Parthership for Soft Condensed Matter Platform (Grenoble, France). A mica disk was cleaned with deionized water and then peeled with sticky tape for use as protein support surface. A volume of 20 µL of several dilution protein samples were deposited on the mica surface for 10 seconds and then, three gently washes of 300 µL of storage buffer (50 mM MES pH 6.0, 500 mM NaCl, 10% Glycerol, 5 mM DTT) were done by directly pipetting on the sample. An extra volume of storage buffer was left on the mica surface for immerse the AFM tip and cantilever. Surfaces were scanned and images acquired by tapping-mode in liquid with an AFM CYPHER S (Asylum Research, Oxford Instruments) using a SNL-C cantilever (Bruker). Cantilever oscillation frequency was tuned to the resonance frequency of ~12.7 kHz and images of 501.96 nm x 611.96 nm were obtained at a scanning frequency of 2.5 Hz. AFM image analysis were done using the visualization and analysis package Gwiddion 2.48 (Nečas *et al*. 2012).

6.6 Analytical ultracentrigufation

Analytical ultracentrifugation experiments were performed using a Beckman Optima XL-I at 20 ºC, operating at velocity sedimentation (SV) mode using a AN-50 Ti rotor with two-channel charcoal-filled centerpieces. Prior to the measurements, NS5 stabilization buffer was exchanged to 50 mM HEPEs pH

7.0, 500 mM NaCl, 5% Glycerol, 0.4 mM DTT using Amicon Ultra 30K centrifugal filters (Millipore). Sedimentation experiments were performed at 43,000 rpm loading protein at 0.5 mg/mL. Data was collected at 280 nm in continuous mode and fitted with SEDFIT.

6.7 SEC with multi-angle laser light scattering

For analysis of the molecular weight of the monomeric and dimeric forms of ZIKV NS5 in solution, SEC combined with light scattering analysis was used. The SEC data was measured using a Superdex 200 10/300 GL column (GE Healthcare; MW range: 10–600 kDa), connected to high-performance liquid chromatography system (Shimadzu), equipped with an autosampler. The elution from SEC was monitored by a UV detector, differential refractometer (OPTI-rEx, Wyatt Corp.) and static, multiangle laser LS detector (DAWN-HELEOS, Wyatt Corp.). The SEC-UV/MALS/RI system was equilibrated in 50 mM MES pH 6.0, 500 mM NaCl, 10% glycerol, 5 mM DTT or 50 mM HEPES pH 7.0, 500 mM NaCl, 10% glycerol, 5 mM DTT at the flow rate of 0.5 mL/min. The ASTRA 7 software (Wyatt Corp.) was used for data collection and analysis.

6.8 NS5 Polymerase activity assay of NS5 Y25A/K28S/K29A

NS5 polymerase elongation activity was measured using a non-radioactive assay as described (Lu *et al.* 2017). Briefly, 1μM of fluorescently labeled RNA oligonucleotide Cy5.5-5'AGAACCUGUUGAACAAAAGC3' was mixed in equal amount with the unlabeled RNA template oligonucleotide 5'AUUAUUCGGCUUUUGUUCAACAGGUUCU3' (IDT, Illinois, USA) in 50 mM NaCl in RNase-free water. RNA mixture was anealed by incubating at 95 ºC for 10 min and slowly cooling to room temperature. Primer extension experiments were carried out in 20 μL reaction containing reaction buffer (5 mM Tris-HCl pH 7.5, 10 mM DTT, 5 mM $MnCl_2$, 0.5% Triton X-100, 0.01 U RNasin, 10% Glycerol), 10 nM primer/template RNA complex and decreasing concentration of NS5 protein. The reaction was initiated by the addition of rNTPs at a final concentration of 100 μM, followed by incubation for 1 h at 35 ºC and then quenched by the addition of 40 μL of quenching buffer (8M Urea, 90 mM Tris base, 29 mM Taurine, 10 mM EDT, 0.02% SDS, 0.1% Bromophenol Blue). Samples were denatured at 95 ºC for 15 min, and the primer extension products were resolved on 15%

denaturing polyacrylamide gels Urea-PAGE in 1X TBE buffer. Gels were visualized using an Odyssey infrared imaging system and images were analyzed using Fiji software.

6.9 PCR-introduced mutagenesis

PCR-introduced mutagenesis was performed using two approaches. In the first approach, a pair of overlapping primers carrying the mutation were used to perform a rolling circle PCR amplification, obtaining the mutated plasmid. The second approach was described by Weiner *et al* 1994. Parental DNA plasmid was used as template for a PCR in which only one primer carried the mutation, after amplification, DNA was treated with DpnI and PNK enzymes (Thermo Scientific) and then DNA ends were ligated with T4 ligase (Thermo Scientific) following the manufacturer protocol.

6.10 *In-vitro* RNA transcription and purification

ZIKV 5SLA RNA sequence was obtained from isolate S-542/Yucatan (Access: MT507050.1), secondary structures were modeled using RNAfold server (Gruber *et al.* 2008). 5SLA sequence wasamplified by high fidelity Phusion DNA polymerase using the primers: Fw: 5' TAATACGACTCACTATAGGGTGATCTGTGTGAATCAGACT 3' and Rv 5' GGGTGATACTGTTGCTAGCTTTCGCTTCAAAC 3' that carried at 5' a T7 promotor sequence and appropriate enzyme restriction sites. PCR amplified products were cloned into a SmaI linearized pUC18 vector and transformed into *E. Coli* Strain DH5a. The same method was followed for the rest of 5SLA constructs. Positive clones were detected by PCR using pUC sequencing primers and confirmed by Sanger sequencing. Plasmid DNA was linearized using restriction enzymes in 500µL individual reactions that were incubated overnight at 37 ºC, digested DNA was purified by mixing equal volumes of restriction reaction with Tris stabilized Phenol Chloroform Isoamyl Alcohol mixture (25:24:1) that were vortex gently and centrifuged at room temperature for 5 min at 10000 g. Upper phase was recovered and mixed with an equal volume of Chloroform repeating the same process, upper phase was recovered and precipitated with 1/10th volume of 3M Sodium Acetate pH 5.2 and 2 volumes of ice chilled 100% Ethanol. Samples were kept on ice for 1h or -20 ºC overnight and then centrifuged

at 4 ºC for 45 min at 20000 g. Salt residues were removed by gently wash with 1 ml 70% Ethanol and letting dry at room temperature. Dried DNA pellet was then disolved in 5 mM Tris pH 7.5, quantified using a Biodrop system and stored at -20 ºC until use.

RNA transcription reactions were carried out using in-house purified T7 RNA polymerase as described. *In-vitro* RNA transcription reactions of 10 mL were assembled at room temperature using 50 ng/µL linear template DNA, 5 mM rNTP and 0.5 mg/mL RNA T7 polymerase in 40 mM Tris-HCl pH 7.9, 3 mM Spermidine, 0.01% Triton X-100, 25 mM $MgCl_2$, 15mM DTT in DEPC treated MilliQ water. Reactions were incubated for 3 h at 37 ºC and stored at -20 ºC until use.

RNA was thaw and centrifuged several times at 1000 rpm for 1 min at 4 ºC to remove residual pyrophosphate from the reaction and injected into a MonoQ 5/50 column equilibrated with 50 mM MOPS pH 6.5, 100 mM NaCl, 0,5 mM EDTA in DEPC MilliQ treated water and eluted with a gradient of 50 mM MOPS pH 6.5, 2M NaCl, 0.5 mM EDTA. A volume of 8 µL of fractions that showed an increased absorbance at 260 nm were mixed with an equal volume of 2X RNA loading dye (95% Formamide, 20 mM EDTA, 0,025% Xylene Cyanol, 0,025% Bromophenol Blue), incubated for 1 min at 95 ºC and then loaded in a 8M Urea 0.5 X TBE PAGE, best fractions were mixed together and concentrated with a 10K centrifugal device (Amicon, Millipore). RNA concentration was measured using a BioDrop system and an extinction coefficient at 260 nm of 851100 M^{-1} cm^{-1}, aliquoted RNA was stored at -80 ºC until use. All RNA samples were annealed before use by heating for 5 min at 85 ºC and then cooling down to 4 ºC at 0.1 ºC/sec, samples were kept on ice until use.

6.11 NS5-5SLA complex formation

Previous to complex formation, frozen 5SLA-RNA was thaw on ice and annealed by incubating for 5 min at 85 ºC and lowering the temperature to 4 ºC at 0.1 ºC/s, RNA sample was then kept on ice until use. ZIKV NS5-5SLA RNA complex was assembled by mixing equal molar amounts of concentrated NS5 and 5SLA RNA stocks in 420 µL of buffer 50 mM Bis-Tris Propane pH 8.0, 300 mM KCl, 10 mM DTT, sample buffer was completely exchanged with a 10K centrifugal device (Amicon, Millipore) and was then concentrated until desired concentration.

6.12 Negative Staining Transmission Electron Microscopy

Negative staining grids were prepared using Carbon coated Copper grids 200 mesh that were glow discharged for 30 seconds at 25 mAh and used within the next 30 minutes. All samples were centrifuged for 45 min at 4 ºC at 20.000 g. 1/10 and 1/50 dilutions of NS5 or NS5-5SLA at 3.5 µM (~350 ng/ml) were used for imaging. A volume of 3 µL of sample solution were loaded into the carbon face of the grid and left to adhere for 1 min and 30 seconds. Excess of sample was blotted from the grid with a piece of filter paper and gently washed with three successive 70µL deionized water drops for 10 second each. Excess of water was blotted carefully and 5 µL of 2% uranyl acetate solution was applied to the carbon side leaving to stain for 1 minute and 30 seconds, uranyl acetate excess was then blotted leaving a thin layer of staining agent. Grids were carefully air dried for 30 seconds and stored in a grid box that was kept in a desiccator at least 4 hours before being imaged.

6.13 Preparation of cryo-EM samples and data acquisition

Several grids types were explored such as C-Flat 2/2 grids, Multi Hole cryo-EM grids, Quantifoil Cu/Rh 0.6/1 and Quantifoil Cu/Rh 1.2/1.3. Several dilutions of NS5 5SLA complex were made from 5 µM to 3.5 µM previous to grid vitrification to find an appropriate particle concentration. Final, samples were prepared using a concentration of 3.5µM of NS5 5SLA complex in Quantifoil 1.2/1.3 Cu/Rh 300 mesh grids. Several holey carbon grids were glow discharged for 60 seconds at 25 mA and used within the next 30 min. Using a Vitrobot system previously equilibrated at 4 ºC and 95% relative humidity, a 3 µL drop of protein solution was loaded into the carbon side of the holey carbon grid for 3 seconds and then blotted with a force of -3, grids were then immediately vitrified in liquid nitrogen cooled ethane and placed carefully in grid storage boxes. Before data acquisition, each grid was previously loaded into a 200 kV Talos Arctica microscope (FEI) equipped with a Falcon III direct electron detector for sample quality checking. Several movies were recorded in counting mode using a magnification of 120000X that corresponded to 0.85 A/px, areas that showed best ice thickness and particle concentration were identified for posterior data acquisition.

Cyo-EM data was acquired using a 200 kV Talos Arctica microscope (FEI) equipped with a Falcon III direct electron detector set in counting mode and using

a magnification of 120000X that corresponded to 0.85 A/px. After loading the sample, basic microscope alignments were performed and data acquisition was scheduled to record 41.69 sec exposure movies of 64 frames each at a total dose of 32.10 e/A^2 between defocus values ranging from 1 to 2.5 µM.

6.14 Cryo-EM data processing

A total of 808 movies were used for data processing. 3700 particles were manually picked, extracted with box size of a 200 px and submitted to 2D classification with with 10 classes, soft mask of 170 Å, E-step= 10 and T=2. Obtained averages were used as templates for autopicking with Gautomatch using a threshold of 0.4 and particle diameter of 150 Å. The program picked automatically 204k particles that were submitted to 2D classification with the same parameters except for the number of classes that were set to 50. Larger classes from classification were used to build an initial model and used for subsequent 3D classification with T=2 and 4 classes in RELION (Scheres *et al.* 2012). We obtained a subset of 27k particles that was used for 3D classifications and refinements yielding the final NS5-5SLA cryo-EM map. The NS5 cryo-EM map was obtained by signal subtraction of the 5SLA signal from the NS5-5SLA map particle extraction and subsequent 3D refinement. The NS5 crystal structure was manually fitted inside the subtracted NS5 cryo-EM map using Chimera (Pettersen *et al.,* 2004) and the NS5 MTase and RdRP domains were rigid-body fitted into the density map using Coot. The NS5 model was further divided into three parts containing the MTase (residues from 1 - 260), RdRP finger and palm (residues 268 - 687) and thumb (residues 688 - 888). The priming loop (residues 787 - 810) was manually accomodated occupying an extra density in the cryo-EM map. The loops located between the MTase and RdRP (residues 261 - 267) and between the fingers and thumb subdomains (residues 688 - 714) were located manually too. The crystal NS5 model was rigid body fitted inside the subtracted cryo-EM map and then real space refined in Phenix. The 5SLA cryo-EM map was obtained using masked 3D classification on MTase and 5SLA map regions. A *de novo* predicted model for the 5SLA sequence was obtained using the RNAComposer web server (Antczak *et al.,* 2016) and fitted manually into the 5SLA cryo-EM map using Chimera and used as template for building in Autodrrafter (Kappel *et al.,* 2019). To model the 5SLA molecule, a combination

of *ab-initio* prediction server RNAComposer and the RNA building tool auto-DRRAFTER produced a 5SLA model that fitted the cryo-EM map.

6.15 EMSA assays

Reactions of 10 µL each were assembled at room temperature using a fixed RNA concentration of 0.45 µM and increasing concentrations of purified NS5 protein ranging from (0.9 to 7.2 µM) in buffer 5 mM Bis-Tris Propane pH 8.0, 300 mM KCl, 10 mM DTT, 5% Glycerol. Reactions were incubated for 45 min at 4 ºC before being analyzed. Electrophoretical mobility shift assays were performed on 0.5X TBE 8% continuous polyacrylamide gels using miniprotean electrophoresis system (Bio-rad). Gels were previously run for 1h at 100V at 4ºC in 0,5X TBE buffer before loading samples that were then resolved at a constant voltage of 100V at 4 ºC. A control well was loaded with 10 µL of 2X RNA loading buffer that contained 0,025% Bomophenol Blue and 0,025% Xylene Cyanol to follow the run. Gels were stained with SYBR gold and visualized with a UV transiluminator. Band integration was done using ImageJ (Schneider *et al.* 2012). Data was analyzed, adjusted and plotted using Prism.

6.16 NS5 *de novo* polymerization of homopolymeric ssRNA

De novo polymerization experiments were performed as described in Sáez-Álvarez *et al.* 2019. Reactions carrying buffer, ATP, NS5 and 5SLA were preincubated for 45 min at 4 ºC before adding the PolyU RNA template (~1000 nt). Triplicated reactions of 25 µL were set up carrying wild type NS5 in presence or absence of 5SLA. The NS5 variant carrying a mutation at the active site that is unable to polymerize (NS5 GAD) was used control. Polymerization assays were done using a poly(U) ssRNA template (Merk) and dsRNA production was detected by using 1X SYBR Gold in the reaction. Fluorescence was measured using a Synergy H1M fluorescence plate reader set at excitation and emission wavelengths of 485 and 520 nm respectively. For data analysis, the curves corresponding to NS5 GAD were subtracted for each corresponding sample. Plots were generated using Prism and curve fitting was done with the same software.

6.17 NS5-5SLA analytical size exclusion chromatography

Size exclusion experiments were performed in buffer 50 mM Bis-Tris Propane pH 8.0, 300 mM KCl, 10 mM DTT. NS5 and 5SLA RNA samples were prepared alone or complexed at a 1:1 molar ratio in a final volume of 10µL. Stock and complexed samples were loaded in an Superdex 200 3/30 SEC coupled to an ÄKTA Micro FPLC system previously equilibrated with the same buffer and eluted with one column volume using a constant flow of 0.1 mL/min. Absorbance values at 260 nm and 280 nm were auto-zero before sample injection and recorded along the whole run.

6.18 Thermal Shift Assay or Differential scanning fluorimetry

A matrix of stocks buffers (SPG, MES, Citrate, Bis-Tris Propane, Na/K Phosphate, Hepes, Imidazole Tris) was designed with different pH values as described in Chari A. et al. 2015. An initial master mix buffer of 300 mM KCl was prepared and a 96 well qPCR plate was loaded with 16 µL each well. Five microliter of 500 mM buffer from the buffer matrix was added to each plate well and mix, then, two small drops of 2µL each containing a 20µM protein solution and 62X SYPRO dye (Thermo Scientific) were placed at the opposite tube walls and kept on ice. The plate was centrifuged for 1 minute at 1000 rpm in a JS 5.9 rotor and inmediately placed on a StepOne Plus Real-Time PCR system. A program consisting in a gradient of temperature was performed from 4 to 95 ºC and fluorescence values for ROX were recorded. Melting curves were exported and analyzed using Applied Biosystems Protein Thermal Shift software.

6.19 Hydrogen Deuterium Mass Exchange

Hydrogen Deuterium Mass Exchange **(**HDX-MS) experiments were performed at Biocev, Czech Republic. Samples were properly diluted in 50 mM HEPES pH 7.5, 250 mM KCl, 10% Glycerol, 5 mM DTT, 1 mM $MgCl_2$. The deuterium labeling was initiated by a 10-fold dilution by D2O-based HEPES buffer, the deuteration was left to proceed at 21°C for 0.33, 2, 20 and 180 min. Protein concentration during the exchange was 5 µM. Following the labelling, 50 µl aliquots were quenched by the addition of 50 µl 6M guanidine, 0.9 M TCEP and 1 M glycine pH 2.4, and incubated for 10 min at 0 °C before being rapidly frozen in liquid nitrogen for proper unfolding and digestion efficiency. Upon quick thawing, the quenched

samples were injected into an LC system consisting of immobilized protease column, peptide desalting microtrap (Michrom Bioresources) and analytical chromatographic column Jupiter C18. Proteases used for the digestion were either porcine pepsin, recombinant rhizopuspepsin or recombinant nepenthesin-1. Digestion was carried out in 0.4% formic acid in water. Peptides were separated by a linear gradient of 10- 35% B in 12 min, followed by a jump to 99% B, where solvent A was 0.2% formic acid/2% acetonitrile in water and B was 95% acetonitrile/0.2% formic acid in water. All the fluidic pathway, protease and analytical columns were kept at 0°C to keep the deuterium back-exchange to a minimum. The LC system was directly interfaced to an electrospray ionization source of a 15T FT-ICR mass spectrometer. The instrument was operated either in LC-MS/MS mode (peptide identification) or LC- MS mode (analysis of deuterated samples). Tandem mass spectrum data were searched against a MASCOT database containing the sequence of MtCDH. HDX-MS data were processed using DeutEx, an in-house developed program.

6.20 Dynamic Light Scattering

NS5-5SLA sample solutions were prepared at 1 mg/mL at different salt concentration ranging from 100 to 400 mM KCl in the buffer 50 mM HEPES pH 7.5, 1 mM $MgCl_2$ and 5 mM DTT. A sample volume of 13 µl were gently loaded in a micro cuvette that was equilibrated for 60 min at 10 ºC before being measured using a Zetasizer Nano (Malvern Panalytical). Distribution by volume and intensity plots were generated and analyzed for each group of salt conditions using provider software.

6.21 5SLA crystallization

RNA stock solutions were thaw on ice and annealed as descried before. Buffer was exchanged using 0.5 mL centrifugal filter units (Amicon 10K 0.5ml) to 20 mM Hepes pH 7.5, 10 mM KCl, 10 mM $MgCl_2$ and concentrated to 8 mg/m. Using the commercial available crystallization screenings Helix (Molecular Dimensions) and Natrix (Hampton research), several 96-well crystallization screening plates were setup by mixing 100 nl of RNA solution with 100 nl of crystallization solution using a Crystal Phoenix (Art Robbins Instruments) nanolitre-drop crystallization robot. Plates were incubated at 4ºC and crystal apparition was monitored

regularly. First crystals appeared at 8 mg/ml in the following conditions A) 0,1 M Sodium Chloride, 5 mM Strontium Chloride hexahydrate, 50 mM MES pH 6.5, 25% v/v PEG 400, B) in 0.3 M Sodium Chloride, 50 mM HEPES pH 6.5, 30% v/v MPD, C) in 0,1M Lithium Chloride, 10mM Manganese(II) chloride tatrahydrate, 50 mM MES pH 6.5, 17% w/v PEG 4000, D) in 0,1 M Sodium Chloride, 50 mM Lithium Chloride, 10 mM Magnesium chloride hexahydrate, 50 mM MES pH 6.5, 17% w/v PEG 4000 and E) in 0,1 M Sodium Chloride, 5 mM Strontium Chloride hexahydrate, 50 mM MES pH 6,5, 17% PEG 4000. All crystal producing conditions were optimized, cryo-cooled in liquid nitrogen and stored until diffraction patterns were measured at an X-Ray synchrotron source.

7 Bibliography

Alvarez, D. E., Ezcurra, A. L. D. L., Fucito, S., & Gamarnik, A. V. (2005). Role of RNA structures present at the 3′ UTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology*, *339*(2), 200-212.

Antczak, M., Popenda, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., ... & Szachniuk, M. (2016). New functionality of RNAComposer: application to shape the axis of miR160 precursor structure. *Acta Biochimica Polonica*, *63*(4), 737-744.

Appleby, T. C., Perry, J. K., Murakami, E., Barauskas, O., Feng, J., Cho, A., ... & Sofia, M. J. (2015). Structural basis for RNA replication by the hepatitis C virus polymerase. *Science*, *347*(6223), 771-775.

Arakawa, M., & Morita, E. (2019). Flavivirus replication organelle biogenesis in the endoplasmic reticulum: Comparison with other single-Stranded positive-sense RNA viruses. *International journal of molecular sciences*, *20*(9), 2336.

Arragain, B., Effantin, G., Gerlach, P., Reguera, J., Schoehn, G., Cusack, S., & Malet, H. (2020). Pre-initiation and elongation structures of full-length La Crosse virus polymerase reveal functionally important conformational changes. *bioRxiv*.

Artsob, H., Gubler, D. J., Enria, D. A., Morales, M. A., Pupo, M., Bunning, M. L., & Dudley, J. P. (2009). West Nile Virus in the New World: trends in the spread and proliferation of West Nile Virus in the Western Hemisphere. *Zoonoses and public health*, *56*(6-7), 357-369.

Bai, X. C., McMullan, G., & Scheres, S. H. (2015). How cryo-EM is revolutionizing structural biology. *Trends in biochemical sciences*, *40*(1), 49-57.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological reviews*, *35*(3), 235.

Bardina, S. V., Bunduc, P., Tripathi, S., Duehr, J., Frere, J. J., Brown, J. A., ... & Stramer, S. L. (2017). Enhancement of Zika virus pathogenesis by preexisting antiflavivirus immunity. *Science*, *356*(6334), 175-180.

Barjas-Castro, M. L., Angerami, R. N., Cunha, M. S., Suzuki, A., Nogueira, J. S., Rocco, I. M., ... & Stucchi, R. S. (2016). Probable transfusion-transmitted Zika virus in Brazil. *Transfusion*, *56*(7), 1684-1688.

Beese, L. S., & Steitz, T. A. (1991). Structural basis for the 3′-5′ exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *The EMBO journal*, *10*(1), 25-33.

Besnard, M., Lastere, S., Teissier, A., Cao-Lormeau, V. M., & Musso, D. (2014). Evidence of perinatal transmission of Zika virus, French Polynesia, December 2013 and February 2014. *Eurosurveillance*, *19*(13), 20751.

Brasil, P., Vasconcelos, Z., Kerin, T., Gabaglia, C. R., Ribeiro, I. P., Bonaldo, M., ... & Tsui, I. (2020). Zika virus vertical transmission in children with confirmed antenatal exposure. *Nat Communications*. 2020;11(1):3510

Bredenbeek, P. J., Kooi, E. A., Lindenbach, B., Huijkman, N., Rice, C. M., & Spaan, W. J. (2003). A stable full-length yellow fever virus cDNA clone and the

role of conserved RNA elements in flavivirus replication. *Journal of General Virology*, *84*(5), 1261-1268.

Brooks, Andrew J., et al. "The interdomain region of dengue NS5 protein that binds to the viral helicase NS3 contains independently functional importin β1 and importin α/β-recognized nuclear localization signals." *Journal of Biological Chemistry* 277.39 (2002): 36399-36407.

Broutet, N., Krauer, F., Riesen, M., Khalakdina, A., Almiron, M., Aldighieri, S., ... & Dye, C. (2016). Zika virus as a cause of neurologic disorders. *New England Journal of Medicine*, *374*(16), 1506-1509.

Bujalowski, P. J., Bujalowski, W., & Choi, K. H. (2017). Interactions between the dengue virus polymerase NS5 and stem-loop A. *Journal of virology*, *91*(11).

Bujalowski, P. J., Bujalowski, W., & Choi, K. H. (2020). Identification of the viral RNA promoter stem loop A (SLA)-binding site on Zika virus polymerase NS5. *Scientific Reports*, *10*(1), 1-13.

Bussetta, C., & Choi, K. H. (2012). Dengue virus nonstructural protein 5 adopts multiple conformations in solution. *Biochemistry*, *51*(30), 5921-5931.

Campbell, M. G., Cheng, A., Brilot, A. F., Moeller, A., Lyumkis, D., Veesler, D., ... & Grigorieff, N. (2012). Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure*, *20*(11), 1823-1828.

Campos, G. S., Bandeira, A. C., & Sardi, S. I. (2015). Zika virus outbreak, bahia, brazil. *Emerging infectious diseases*, *21*(10), 1885.

Cao-Lormeau, V. M., Roche, C., Teissier, A., Robin, E., Berry, A. L., Mallet, H. P., ... & Musso, D. (2014). Zika virus, French polynesia, South pacific, 2013. *Emerging infectious diseases*, *20*(6), 1085.

Chari, A., Haselbach, D., Kirves, J. M., Ohmer, J., Paknia, E., Fischer, N., ... & Jarvis, M. (2015). ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nature methods*, *12*(9), 859-865.

Chaudhary, V., Yuen, K. S., Chan, J. F. W., Chan, C. P., Wang, P. H., Cai, J. P., ... & Yuen, K. Y. (2017). Selective activation of type II interferon signaling by Zika virus NS5 protein. *Journal of virology*, *91*(14).

Cheng, Y., Glaeser, R. M., & Nogales, E. (2017). How cryo-EM became so hot. *Cell*, *171*(6), 1229-1231.

Chinnaswamy, S., Murali, A., Li, P., Fujisaki, K., & Kao, C. C. (2010). Regulation of de novo-initiated RNA synthesis in hepatitis C virus RNA-dependent RNA polymerase by intermolecular interactions. *Journal of virology*, *84*(12), 5923-5935.

Choi, K. H., & Rossmann, M. G. (2009). RNA-dependent RNA polymerases from Flaviviridae. *Current opinion in structural biology*, *19*(6), 746-751.

Choi, K. H., Groarke, J. M., Young, D. C., Kuhn, R. J., Smith, J. L., Pevear, D. C., & Rossmann, M. G. (2004). The structure of the RNA-dependent RNA polymerase from bovine viral diarrhea virus establishes the role of GTP in de novo initiation. *Proceedings of the National Academy of Sciences*, *101*(13), 4425-4430.

Clemente-Casares, P., López-Jiménez, A. J., Bellón-Echeverría, I., Encinar, J. A., Martínez-Alfaro, E., Pérez-Flores, R., & Mas, A. (2011). De novo polymerase activity and oligomerization of hepatitis C virus RNA-dependent RNA-polymerases from genotypes 1 to 5. *PLoS One*, *6*(4), e18515.

Clyde, K., & Harris, E. (2006). RNA secondary structure in the coding region of dengue virus type 2 directs translation start codon selection and is required for viral replication. *Journal of virology*, *80*(5), 2170-2182.

Clyde, K., Barrera, J., & Harris, E. (2008). The capsid-coding region hairpin element (cHP) is a critical determinant of dengue virus and West Nile virus RNA synthesis. *Virology*, *379*(2), 314-323.

Clyde, K., Kyle, J. L., & Harris, E. (2006). Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. *Journal of virology*, *80*(23), 11418-11431.

Coloma, J., Jain, R., Rajashankar, K. R., García-Sastre, A., & Aggarwal, A. K. (2016). Structures of NS5 methyltransferase from Zika virus. *Cell reports*, *16*(12), 3097-3102.

Dang, J., Tiwari, S. K., Lichinchi, G., Qin, Y., Patil, V. S., Eroshkin, A. M., & Rana, T. M. (2016). Zika virus depletes neural progenitors in human cerebral organoids through activation of the innate immune receptor TLR3. *Cell stem cell*, *19*(2), 258-265.

De la Peña, M., Kyrieleis, O. J., & Cusack, S. (2007). Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. *The EMBO Journal*, *26*(23), 4913-4925.

Decroly, E., Ferron, F., Lescar, J., & Canard, B. (2012). Conventional and unconventional mechanisms for capping viral mRNA. *Nature Reviews Microbiology*, *10*(1), 51-65.

Den Boon, J. A., & Ahlquist, P. (2010). Organelle-like membrane compartmentalization of positive-strand RNA virus replication factories. *Annual review of microbiology*, *64*, 241-256.

Deng, B., Lento, C., & Wilson, D. J. (2016). Hydrogen deuterium exchange mass spectrometry in biopharmaceutical discovery and development–A review. *Analytica chimica acta*, *940*, 8-20.

Deng, R., & Brock, K. V. (1993). 5′ and 3′ untranslated regions of pestivirus genome: primary and secondary structure analyses. *Nucleic acids research*, *21*(8), 1949-1957.

Deng, Y. Q., Zhao, H., Li, X. F., Zhang, N. N., Liu, Z. Y., Jiang, T., ... & Sun, Z. Z. (2016). Isolation, identification and genomic characterization of the Asian lineage Zika virus imported to China. *Science China Life Sciences*, *59*(4), 428-430.

Dick, G. W. A., Kitchen, S. F., & Haddow, A. J. (1952). Zika virus (I). Isolations and serological specificity. *Transactions of the royal society of tropical medicine and hygiene*, *46*(5), 509-520.

Dong, H., Fink, K., Züst, R., Lim, S. P., Qin, C. F., & Shi, P. Y. (2014). Flavivirus RNA methylation. *Journal of General Virology*, *95*(4), 763-778.

Dong, H., Zhang, B., & Shi, P. Y. (2008). Terminal structures of West Nile virus genomic RNA and their interactions with viral NS5 protein. *Virology*, *381*(1), 123-135.

Duffy, M. R., Chen, T. H., Hancock, W. T., Powers, A. M., Kool, J. L., Lanciotti, R. S., ... & Guillaumot, L. (2009). Zika virus outbreak on Yap Island, federated states of Micronesia. *New England Journal of Medicine*, *360*(24), 2536-2543.

Edgil, D., Polacek, C., & Harris, E. (2006). Dengue virus utilizes a novel strategy for translation initiation when cap-dependent translation is inhibited. *Journal of virology*, *80*(6), 2976-2986.

Egloff, M. P., Benarroch, D., Selisko, B., Romette, J. L., & Canard, B. (2002). An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *The EMBO journal*, *21*(11), 2757-2768.

El Sahili, A., & Lescar, J. (2017). Dengue virus non-structural protein 5. *Viruses*, *9*(4), 91.

Fernandez-Leiro, R., & Scheres, S. H. (2016). Unravelling biological macromolecules with cryo-electron microscopy. *Nature*, *537*(7620), 339-346.

Ferrer-Orta, C., Arias, A., Escarmís, C., & Verdaguer, N. (2006). A comparison of viral RNA-dependent RNA polymerases. *Current opinion in structural biology*, *16*(1), 27-34.

Ferrero, D. S., Buxaderas, M., Rodríguez, J. F., & Verdaguer, N. (2015). The structure of the RNA-dependent RNA polymerase of a Permutotetravirus suggests a link between primer-dependent and primer-independent polymerases. *PLoS pathogens*, *11*(12), e1005265.

Ferrero, D., Ferrer-Orta, C., & Verdaguer, N. (2018). Viral RNA-dependent RNA polymerases: a structural overview. In *Virus Protein and Nucleoprotein Complexes* (pp. 39-71). Springer, Singapore.

Filomatori, C. V., Lodeiro, M. F., Alvarez, D. E., Samsa, M. M., Pietrasanta, L., & Gamarnik, A. V. (2006). A 5′ RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes & development*, *20*(16), 2238-2249.

Filomatori, C. V., Lodeiro, M. F., Alvarez, D. E., Samsa, M. M., Pietrasanta, L., & Gamarnik, A. V. (2006). A 5′ RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes & development*, *20*(16), 2238-2249.

Fourcade, C., Mansuy, J. M., Dutertre, M., Delpech, M., Marchou, B., Delobel, P., ... & Martin-Blondel, G. (2016). Viral load kinetics of Zika virus in plasma, urine and saliva in a couple returning from Martinique, French West Indies. *Journal of Clinical Virology*, *82*, 1-4.

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., ... & Svergun, D. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography*, *50*(4), 1212-1225.

Friebe, P., & Harris, E. (2010). Interplay of RNA elements in the dengue virus 5′ and 3′ ends required for viral RNA replication. *Journal of virology*, *84*(12), 6103-6118.

Fung, J. (2015). Era of direct acting antivirals in chronic hepatitis C: Who will benefit?. *World journal of hepatology*, *7*(24), 2543.

Gabriel, E., Ramani, A., Karow, U., Gottardo, M., Natarajan, K., Gooi, L. M., ... & Kuivanen, S. (2017). Recent Zika virus isolates induce premature differentiation of neural progenitors in human brain organoids. *Cell stem cell*, *20*(3), 397-406.

Garcez, P. P., Loiola, E. C., da Costa, R. M., Higa, L. M., Trindade, P., Delvecchio, R., ... & Rehen, S. K. (2016). Zika virus impairs growth in human neurospheres and brain organoids. *Science*, *352*(6287), 816-818.

Gardner, C. L., & Ryman, K. D. (2010). Yellow fever: a reemerging threat. *Clinics in laboratory medicine*, *30*(1), 237-260.

Ghaffar, K. A., Ng, L. F., & Renia, L. (2018). Fast tracks and roadblocks for zika vaccines. *Vaccines*, *6*(4), 77.

Gong, P., & Peersen, O. B. (2010). Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proceedings of the National Academy of Sciences*, *107*(52), 22505-22510.

Gorbalenya, A. E., Pringle, F. M., Zeddam, J. L., Luke, B. T., Cameron, C. E., Kalmakoff, J., ... & Ward, V. K. (2002). The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *Journal of molecular biology*, *324*(1), 47-62.

Govero, J., Esakky, P., Scheaffer, S. M., Fernandez, E., Drury, A., Platt, D. J., ... & Moley, K. H. (2016). Zika virus infection damages the testes in mice. *Nature*, *540*(7633), 438-442.

Grant, A., Ponia, S. S., Tripathi, S., Balasubramaniam, V., Miorin, L., Sourisseau, M., ... & García-Sastre, A. (2016). Zika virus targets human STAT2 to inhibit type I interferon signaling. *Cell host & microbe*, *19*(6), 882-890.

Greene, N. D., & Copp, A. J. (2009). Development of the vertebrate central nervous system: formation of the neural tube. *Prenatal Diagnosis: Published in Affiliation with the International Society for Prenatal Diagnosis*, *29*(4), 303-311.

Gritsun, T. S., & Gould, E. A. (2006). Origin and evolution of 3′ UTR of flaviviruses: long direct repeats as a basis for the formation of secondary structures and their significance for virus transmission. *Advances in virus research*, *69*, 203-248.

Gubler, D. J. (2002). The global emergence/resurgence of arboviral diseases as public health problems. *Archives of medical research*, *33*(4), 330-342.

Gubler, D. J., & Roehrig, J. T. (1998). Arboviruses (Togaviridae and Flaviviridae). *Topley and Wilson's microbiology and microbial infections*, *1*, 579-600.

Hahn, C. S., Hahn, Y. S., Rice, C. M., Lee, E., Dalgarno, L., Strauss, E. G., & Strauss, J. H. (1987). Conserved elements in the 3′ untranslated region of flavivirus RNAs and potential cyclization sequences. *Journal of molecular biology*, *198*(1), 33-41.

Hamel, R., Dejarnac, O., Wichit, S., Ekchariyawat, P., Neyret, A., Luplertlop, N., ... & Cao-Lormeau, V. M. (2015). Biology of Zika virus infection in human skin cells. *Journal of virology*, *89*(17), 8880-8896.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics*, *28*(2), 171-193.

Herzik, M. A., Wu, M., & Lander, G. C. (2019). High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nature communications*, *10*(1), 1-9.

Heymann, D. L., Hodgson, A., Freedman, D. O., Staples, J. E., Althabe, F., Baruah, K., ... & Menon, K. U. (2016). Zika virus and microcephaly: why is this situation a PHEIC?. *The Lancet*, *387*(10020), 719-721.

Hobson, S. D., Rosenblum, E. S., Richards, O. C., Richmond, K., Kirkegaard, K., & Schultz, S. C. (2001). Oligomeric structures of poliovirus polymerase are important for function. *The EMBO journal*, *20*(5), 1153-1163.

Hodge, K., Tunghirun, C., Kamkaew, M., Limjindaporn, T., Yenchitsomanus, P. T., & Chimnaronk, S. (2016). Identification of a conserved RNA-dependent RNA polymerase (RdRp)-RNA interface required for flaviviral replication. *Journal of Biological Chemistry*, *291*(33), 17437-17449.

Högbom, M., Jäger, K., Robel, I., Unge, T., & Rohayem, J. (2009). The active form of the norovirus RNA-dependent RNA polymerase is a homodimer with cooperative activity. *Journal of General Virology*, *90*(2), 281-291.

ICTV: https://ictv.global/taxonomy/

Issur, M., Geiss, B. J., Bougie, I., Picard-Jean, F., Despins, S., Mayette, J., ... & Bisaillon, M. (2009). The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. *Rna*, *15*(12), 2340-2350.

Jiang, W., Baker, M. L., Jakana, J., Weigele, P. R., King, J., & Chiu, W. (2008). Backbone structure of the infectious ε15 virus capsid revealed by electron cryomicroscopy. *Nature*, *451*(7182), 1130-1134.

Johansson, M., Brooks, A. J., Jans, D. A., & Vasudevan, S. G. (2001). A small region of the dengue virus-encoded RNA-dependent RNA polymerase, NS5, confers interaction with both the nuclear transport receptor importin-β and the viral helicase, NS3. *Journal of General Virology*, *82*(4), 735-745.

Johnson, A. E., & van Waes, M. A. (1999). The translocon: a dynamic gateway at the ER membrane. *Annual review of cell and developmental biology*, *15*(1), 799-842.

Julander, J. G., Siddharthan, V., Evans, J., Taylor, R., Tolbert, K., Apuli, C., ... & Van Wettere, A. (2017). Efficacy of the broad-spectrum antiviral compound BCX4430 against Zika virus in cell culture and in a mouse model. *Antiviral research*, *137*, 14-22.

Kadek, A., Kavan, D., Felice, A. K., Ludwig, R., Halada, P., & Man, P. (2015). Structural insight into the calcium ion modulated interdomain electron transfer in cellobiose dehydrogenase. *FEBS letters*, *589*(11), 1194-1199.

Kamiyama, N., Soma, R., Hidano, S., Watanabe, K., Umekita, H., Fukuda, C., ... & Sachi, N. (2017). Ribavirin inhibits Zika virus (ZIKV) replication in vitro and suppresses viremia in ZIKV-infected STAT1-deficient mice. *Antiviral research*, *146*, 1-11.

Kao, C. C., Singh, P., & Ecker, D. J. (2001). De novo initiation of viral RNA-dependent RNA synthesis. *Virology*, *287*(2), 251-260.

Kappel, K., Zhang, K., Su, Z., Kladwang, W., Li, S., Pintilie, G., ... & Yesselman, J. D. (2019). Ribosolve: rapid determination of three-dimensional RNA-only structures. *BioRxiv*, 717801.

Khromykh, A. A., Meka, H., Guyatt, K. J., & Westaway, E. G. (2001). Essential role of cyclization sequences in flavivirus RNA replication. *Journal of virology*, *75*(14), 6719-6728.

Kidmose, R. T., Vasiliev, N. N., Chetverin, A. B., Andersen, G. R., & Knudsen, C. R. (2010). Structure of the Qβ replicase, an RNA-dependent RNA polymerase consisting of viral and host proteins. *Proceedings of the National Academy of Sciences*, *107*(24), 10884-10889.

Klema, V. J., Padmanabhan, R., & Choi, K. H. (2015). Flaviviral replication complex: coordination between RNA synthesis and 5'-RNA capping. *Viruses*, *7*(8), 4640-4656.

Klema, V. J., Ye, M., Hindupur, A., Teramoto, T., Gottipati, K., Padmanabhan, R., & Choi, K. H. (2016). Dengue virus nonstructural protein 5 (NS5) assembles into a dimer with a unique methyltransferase and polymerase interface. *PLoS pathogens*, *12*(2), e1005451.

Kofler, R. M., Hoenninger, V. M., Thurner, C., & Mandl, C. W. (2006). Functional analysis of the tick-borne encephalitis virus cyclization elements indicates major differences between mosquito-borne and tick-borne flaviviruses. *Journal of virology*, *80*(8), 4099-4113.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of applied crystallography*, *36*(5), 1277-1282.

Kouba, T., Drncova, P., & Cusack, S. (2019). Structural snapshots of actively transcribing influenza polymerase. *Nature structural & molecular biology*, *26*(6), 460-470.

Kriegstein, A., & Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. *Annual review of neuroscience*, *32*, 149–184.

Kriegstein, A., & Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. *Annual review of neuroscience*, *32*, 149-184.

Kumar, A., Hou, S., Airo, A. M., Limonta, D., Mancinelli, V., Branton, W., ... & Hobman, T. C. (2016). Zika virus inhibits type-I interferon production and downstream signaling. *EMBO reports*, *17*(12), 1766-1775.

Leal, M. C., Muniz, L. F., Ferreira, T. S., Santos, C. M., Almeida, L. C., Linden, V. V. D., ... & Neto, S. S. C. (2016). Hearing loss in infants with microcephaly and evidence of congenital Zika virus infection—Brazil, November 2015–May 2016. *Morbidity and Mortality Weekly Report*, *65*(34), 917-919.

Lescar, J., & Canard, B. (2009). RNA-dependent RNA polymerases from flaviviruses and Picornaviridae. *Current opinion in structural biology*, *19*(6), 759-767.

Li, H., Saucedo-Cuevas, L., Regla-Nava, J. A., Chai, G., Sheets, N., Tang, W., ... & Gleeson, J. G. (2016). Zika virus infects neural progenitors in the adult mouse brain and alters proliferation. *Cell stem cell*, *19*(5), 593-598.

Lim, S. P., Noble, C. G., Seh, C. C., Soh, T. S., El Sahili, A., Chan, G. K. Y., ... & Manjunatha, U. (2016). Potent allosteric dengue virus NS5 polymerase inhibitors: mechanism of action and resistance profiling. *PLoS pathogens*, *12*(8), e1005737.

Lin, S., Yang, S., He, J., Guest, J. D., Ma, Z., Yang, L., ... & Zhang, Y. J. (2019). Zika virus NS5 protein antagonizes type I interferon production via blocking TBK1 activation. *Virology*, *527*, 180-187.

Lindenbach, B. D., & Rice, C. M. (1999). Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function. *Journal of virology*, *73*(6), 4611-4621.

Lo, M. K., Tilgner, M., Bernard, K. A., & Shi, P. Y. (2003). Functional analysis of mosquito-borne flavivirus conserved sequence elements within 3′ untranslated region of West Nile virus by use of a reporting replicon that differentiates between viral translation and RNA replication. *Journal of virology*, *77*(18), 10004-10014.

Lodeiro, M. F., Filomatori, C. V., & Gamarnik, A. V. (2009). Structural and functional studies of the promoter element for dengue virus RNA replication. *Journal of virology*, *83*(2), 993-1008.

Lu, G., & Gong, P. (2013). Crystal structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. *PLoS Pathog*, *9*(8), e1003549.

Lu, G., Bluemling, G. R., Collop, P., Hager, M., Kuiper, D., Gurale, B. P., ... & Kolykhalov, A. A. (2017). Analysis of ribonucleotide 5′-triphosphate analogs as potential inhibitors of Zika virus RNA-dependent RNA polymerase by using nonradioactive polymerase assays. *Antimicrobial agents and chemotherapy*, *61*(3).

Luo, D., Xu, T., Watson, R. P., Scherer-Becker, D., Sampath, A., Jahnke, W., ... & Vasudevan, S. G. (2008). Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *The EMBO journal*, *27*(23), 3209-3219.

Ma, L., Jones, C. T., Groesch, T. D., Kuhn, R. J., & Post, C. B. (2004). Solution structure of dengue virus capsid protein reveals another fold. *Proceedings of the National Academy of Sciences*, *101*(10), 3414-3419.

Mackenzie, J. S., Gubler, D. J., & Petersen, L. R. (2004). Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. *Nature medicine*, *10*(12), S98-S109.

Malet, H., Massé, N., Selisko, B., Romette, J. L., Alvarez, K., Guillemot, J. C., ... & Canard, B. (2008). The flavivirus polymerase as a target for drug discovery. *Antiviral research*, *80*(1), 23-35.

Manzano, M., Reichert, E. D., Polo, S., Falgout, B., Kasprzak, W., Shapiro, B. A., & Padmanabhan, R. (2011). Identification of cis-acting elements in the 3′-untranslated region of the dengue virus type 2 RNA that modulate translation and replication. *Journal of Biological Chemistry*, *286*(25), 22521-22534.

Marrs, C., Olson, G., Saade, G., Hankins, G., Wen, T., Patel, J., & Weaver, S. (2016). Zika virus and pregnancy: a review of the literature and clinical considerations. *American journal of perinatology*, *33*(7), 625.

Martínez-Cerdeño, V., & Noctor, S. C. (2018). Neural progenitor cell terminology. *Frontiers in Neuroanatomy*, *12*, 104.

McMullan, G., Faruqi, A. R., Clare, D., & Henderson, R. (2014). Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy*, *147*, 156-163.

Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M. I., Pragani, R., Boxer, M. B., Earl, L. A., Milne, J., & Subramaniam, S. (2016). Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell*, *165*(7), 1698–1707. https://doi.org/10.1016/j.cell.2016.05.040

Mosley, R. T., Edwards, T. E., Murakami, E., Lam, A. M., Grice, R. L., Du, J., ... & Otto, M. J. (2012). Structure of hepatitis C virus polymerase in complex with primer-template RNA. *Journal of virology*, *86*(12), 6503-6511.

Mukhopadhyay, S., Kuhn, R. J., & Rossmann, M. G. (2005). A structural perspective of the flavivirus life cycle. *Nature Reviews Microbiology*, *3*(1), 13-22.

Musso, D., Nhan, T., Robin, E., Roche, C., Bierlaire, D., Zisou, K., ... & Broult, J. (2014). Potential for Zika virus transmission through blood transfusion demonstrated during an outbreak in French Polynesia, November 2013 to February 2014. *Eurosurveillance*, *19*(14), 20761.

Nečas, D., & Klapetek, P. (2012). Gwyddion: an open-source software for SPM data analysis. *Open Physics*, *10*(1), 181-188.

Nguyen, H. N., Qian, X., Song, H., & Ming, G. L. (2016). Neural stem cells attacked by Zika virus. *Cell research*, *26*(7), 753-754.

Nguyen, T. H. D., Galej, W. P., Bai, X. C., Oubridge, C., Newman, A. J., Scheres, S. H., & Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6. U5 tri-snRNP at 3.7 Å resolution. *Nature*, *530*(7590), 298-302.

Olsthoorn, R. C., & Bol, J. F. (2001). Sequence comparison and secondary structure analysis of the 3'noncoding region of flavivirus genomes reveals multiple pseudoknots. *Rna*, *7*(10), 1370-1377.

Pagani, I., Ghezzi, S., Ulisse, A., Rubio, A., Turrini, F., Garavaglia, E., ... & Broccoli, V. (2017). Human endometrial stromal cells are highly permissive to productive infection by Zika virus. *Scientific reports*, *7*, 44286.

Panayiotou, C., Lindqvist, R., Kurhade, C., Vonderstein, K., Pasto, J., Edlund, K., ... & Överby, A. K. (2018). Viperin restricts Zika virus and tick-borne encephalitis virus replication by targeting NS3 for proteasomal degradation. *Journal of virology*, *92*(7).

Patkar, C. G., & Kuhn, R. J. (2008). Yellow fever virus NS3 plays an essential role in virus assembly independent of its known enzymatic functions. *Journal of virology*, *82*(7), 3342-3352.

Paul, D., & Bartenschlager, R. (2013). Architecture and biogenesis of plus-strand RNA virus replication factories. *World journal of virology*, *2*(2), 32.

Paz-Bailey, G., Rosenberg, E. S., Doyle, K., Munoz-Jordan, J., Santiago, G. A., Klein, L., ... & Lozier, M. J. (2018). Persistence of Zika virus in body fluids. *New England Journal of Medicine*, *379*(13), 1234-1243.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605-1612.

Pierson, T. C., Fremont, D. H., Kuhn, R. J., & Diamond, M. S. (2008). Structural insights into the mechanisms of antibody-mediated neutralization of flavivirus infection: implications for vaccine development. *Cell host & microbe*, *4*(3), 229-238.

Pravda, L., Sehnal, D., Toušek, D., Navrátilová, V., Bazgier, V., Berka, K., ... & Otyepka, M. (2018). MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic acids research*, *46*(W1), W368-W373.

Polacek, C., Foley, J. E., & Harris, E. (2009). Conformational changes in the solution structure of the dengue virus 5′ end in the presence and absence of the 3′ untranslated region. *Journal of virology*, *83*(2), 1161-1166.

Potisopon, S., Priet, S., Collet, A., Decroly, E., Canard, B., & Selisko, B. (2014). The methyltransferase domain of dengue virus protein NS5 ensures efficient RNA synthesis initiation and elongation by the polymerase domain. *Nucleic acids research*, *42*(18), 11642-11656.

Renaud, J. P., Chari, A., Ciferri, C., Liu, W. T., Remigy, H. W., Stark, H., & Wiesmann, C. (2018). Cryo-EM in drug discovery: achievements, limitations and prospects. *Nature reviews Drug discovery*, *17*(7), 471-492.

Rice, C. M., Lenches, E. M., Shin, S. J., Sheets, R. L., & Strauss, J. H. (1985). Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science*, *229*(4715), 726-733.

Roby, J. A., Setoh, Y. X., Hall, R. A., & Khromykh, A. A. (2015). Post-translational regulation and modifications of flavivirus structural proteins. *Journal of General Virology*, *96*(7), 1551-1569.

Romero, T. A., Tumban, E., Jun, J., Lott, W. B., & Hanley, K. A. (2006). Secondary structure of dengue virus type 4 3′ untranslated region: impact of deletion and substitution mutations. *Journal of general virology*, *87*(11), 3291-3296.

Rusanov, T., Kent, T., Saeed, M., Hoang, T. M., Thomas, C., Rice, C. M., & Pomerantz, R. T. (2018). Identification of a Small Interface between the Methyltransferase and RNA Polymerase of NS5 that is Essential for Zika Virus Replication. *Scientific reports*, *8*(1), 1-11.

Saade, M., Blanco-Ameijeiras, J., Gonzalez-Gobartt, E., & Martí, E. (2018). A centrosomal view of CNS growth. *Development*, *145*(21).

Saade, M., Ferrero, D. S., Blanco-Ameijeiras, J., Gonzalez-Gobartt, E., Ruiz-Arroyo, V. M., Martínez-Sáez, E., ... & Martí, E. (2019). Multimerization of Zika Virus-NS5 causes a ciliopathy and forces premature neurogenesis. *bioRxiv*, 719625.

Saade, M., Gutiérrez-Vallejo, I., Le Dréau, G., Rabadán, M. A., Miguez, D. G., Buceta, J., & Martí, E. (2013). Sonic hedgehog signaling switches the mode of division in the developing nervous system. *Cell reports*, *4*(3), 492-503.

Sacramento, C. Q., De Melo, G. R., De Freitas, C. S., Rocha, N., Hoelz, L. V. B., Miranda, M., ... & Abrantes, J. L. (2017). The clinically approved antiviral drug sofosbuvir inhibits Zika virus replication. *Scientific reports*, *7*, 40920.

Saeedi, B. J., & Geiss, B. J. (2013). Regulation of flavivirus RNA synthesis and capping. *Wiley Interdisciplinary Reviews: RNA*, *4*(6), 723-735.

Sáez-Álvarez, Y., Arias, A., del Águila, C., & Agudo, R. (2019). Development of a fluorescence-based method for the rapid determination of Zika virus polymerase activity and the screening of antiviral drugs. *Scientific reports*, *9*(1), 1-11.

Saw, W. G., Pan, A., Manimekalai, M. S. S., & Grüber, G. (2017). Structural features of Zika virus non-structural proteins 3 and-5 and its individual domains in solution as well as insights into NS3 inhibition. *Antiviral research*, *141*, 73-90.

Saw, W. G., Tria, G., Grüber, A., Manimekalai, S., Zhao, Y., Chandramohan, A., ... & Grüber, G. (2015). Structural insight and flexible features of NS5 proteins from all four serotypes of Dengue virus in solution. *Acta Crystallographica Section D: Biological Crystallography*, *71*(11), 2309-2327.

Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology*, *180*(3), 519-530.

Seidelt, B., Innis, C. A., Wilson, D. N., Gartmann, M., Armache, J. P., Villa, E., ... & Steitz, T. A. (2009). Structural insight into nascent polypeptide chain–mediated translational stalling. *Science*, *326*(5958), 1412-1415.

Selisko, B., Potisopon, S., Agred, R., Priet, S., Varlet, I., Thillier, Y., ... & Canard, B. (2012). Molecular basis for nucleotide conservation at the ends of the dengue virus genome. *PLoS Pathog*, *8*(9), e1002912.

Selisko, B., Wang, C., Harris, E., & Canard, B. (2014). Regulation of Flavivirus RNA synthesis and replication. *Current opinion in virology*, *9*, 74-83.

Sharp, T. M., Muñoz-Jordán, J., Perez-Padilla, J., Bello-Pagán, M. I., Rivera, A., Pastula, D. M., ... & Waterman, S. (2016). Zika virus infection associated with severe thrombocytopenia. *Clinical Infectious Diseases*, *63*(9), 1198-1201.

Shi, P. Y. (2014). Unraveling a flavivirus enigma. *Science*, *343*(6173), 849-850.

Shiryaev, S. A., Farhy, C., Pinto, A., Huang, C. T., Simonetti, N., Ngono, A. E., ... & Strongin, A. Y. (2017). Characterization of the Zika virus two-component NS2B-NS3 protease and structure-assisted identification of allosteric small-molecule antagonists. *Antiviral research*, *143*, 218-229.

Shu, B., & Gong, P. (2016). Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *Proceedings of the National Academy of Sciences*, *113*(28), E4005-E4014.

Smallwood, S., Hövel, T., Neubert, W. J., & Moyer, S. A. (2002). Different substitutions at conserved amino acids in domains II and III in the Sendai L RNA polymerase protein inactivate viral RNA synthesis. *Virology*, *304*(1), 135-145.

Soares, C. N., Brasil, P., Carrera, R. M., Sequeira, P., De Filippis, A. B., Borges, V. A., ... & Solomon, T. (2016). Fatal encephalitis associated with Zika virus infection in an adult. *Journal of Clinical Virology*, *83*, 63-65.

Solomon, T., Ni, H., Beasley, D. W., Ekkelenkamp, M., Cardosa, M. J., & Barrett, A. D. (2003). Origin and evolution of Japanese encephalitis virus in southeast Asia. *Journal of virology*, *77*(5), 3091-3098.

Stadler, K., Allison, S. L., Schalich, J., & Heinz, F. X. (1997). Proteolytic activation of tick-borne encephalitis virus by furin. *Journal of virology*, *71*(11), 8475-8481.

Sun, J., Wu, D., Zhong, H., Guan, D., Zhang, H., Tan, Q., & Ke, C. (2016). Presence of Zika virus in conjunctival fluid. *JAMA ophthalmology*, *134*(11), 1330-1332.

Sutton, G., Grimes, J. M., Stuart, D. I., & Roy, P. (2007). Bluetongue virus VP4 is an RNA-capping assembly line. *Nature structural & molecular biology*, *14*(5), 449-451.

Svergun, D. I., Petoukhov, M. V., & Koch, M. H. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophysical journal*, *80*(6), 2946-2953.

Takegami, T., Washizu, M., & Yasui, K. (1986). Nucleotide sequence at the 3′ end of Japanese encephalitis virus genomic RNA. *Virology*, *152*(2), 483-486.

Tang, H., Hammack, C., Ogden, S. C., Wen, Z., Qian, X., Li, Y., ... & Christian, K. M. (2016). Zika virus infects human cortical neural progenitors and attenuates their growth. *Cell stem cell*, *18*(5), 587-590.

Tao, Y., Farsetta, D. L., Nibert, M. L., & Harrison, S. C. (2002). RNA synthesis in a cage—structural studies of reovirus polymerase λ3. *Cell*, *111*(5), 733-745.

Tay, M. Y., Saw, W. G., Zhao, Y., Chan, K. W., Singh, D., Chong, Y., ... & Luo, D. (2015). The C-terminal 50 amino acid residues of dengue NS3 protein are important for NS3-NS5 interaction and viral replication. *Journal of Biological Chemistry*, *290*(4), 2379-2394.

te Velthuis, A. J. (2014). Common and unique features of viral RNA-dependent polymerases. *Cellular and molecular life sciences*, *71*(22), 4403-4420.

te Velthuis, A. J., & Fodor, E. (2016). Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nature Reviews Microbiology*, *14*(8), 479-493.

Tickle, I. J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., & Bricogne, G. (2018). Staraniso. *Global Phasing Ltd., Cambridge, UK*.

Tortorici, M. A., & Veesler, D. (2019). Structural insights into coronavirus entry. In *Advances in virus research* (Vol. 105, pp. 93-116). Academic Press.

Upadhyay, A. K., Cyr, M., Longenecker, K., Tripathi, R., Sun, C., & Kempf, D. J. (2017). Crystal structure of full-length Zika virus NS5 protein reveals a conformation similar to Japanese encephalitis virus NS5. *Acta Crystallographica Section F: Structural Biology Communications*, *73*(3), 116-122.

van den Pol, A. N., Mao, G., Yang, Y., Ornaghi, S., & Davis, J. N. (2017). Zika virus targeting in the developing brain. *Journal of Neuroscience*, *37*(8), 2161-2175.

van Dijk, A. A., Makeyev, E. V., & Bamford, D. H. (2004). Initiation of viral RNA-dependent RNA polymerization. *Journal of general virology*, *85*(5), 1077-1093.

Venkataraman, S., Prasad, B. V., & Selvarajan, R. (2018). RNA dependent RNA polymerases: insights from structure, function and evolution. *Viruses*, *10*(2), 76.

Villa, E., Sengupta, J., Trabuco, L. G., LeBarron, J., Baxter, W. T., Shaikh, T. R., ... & Frank, J. (2009). Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proceedings of the National Academy of Sciences*, *106*(4), 1063-1068.

Wandzik, J. M., Kouba, T., Karuppasamy, M., Pflug, A., Drncova, P., Provaznik, J., ... & Cusack, S. (2020). A Structure-Based Model for the Complete Transcription Cycle of Influenza Polymerase. *Cell*.

Warren, T. K., Wells, J., Panchal, R. G., Stuthman, K. S., Garza, N. L., Van Tongeren, S. A., ... & Honnold, S. (2014). Protection against filovirus diseases by a novel broad-spectrum nucleoside analogue BCX4430. *Nature*, *508*(7496), 402-405.

Weiner, M. P., Costa, G. L., Schoettlin, W., Cline, J., Mathur, E., & Bauer, J. C. (1994). Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene*, *151*(1-2), 119-123.

Welsch, S., Miller, S., Romero-Brey, I., Merz, A., Bleck, C. K., Walther, P., ... & Bartenschlager, R. (2009). Composition and three-dimensional architecture of the dengue virus replication and assembly sites. *Cell host & microbe*, *5*(4), 365-375.

Wu, J., Lu, G., Zhang, B., & Gong, P. (2015). Perturbation in the conserved methyltransferase-polymerase interface of flavivirus NS5 differentially affects polymerase initiation and elongation. *Journal of virology*, *89*(1), 249-261.

Xie, X., Gayen, S., Kang, C., Yuan, Z., & Shi, P. Y. (2013). Membrane topology and function of dengue virus NS2A protein. *Journal of virology*, *87*(8), 4609-4622.

Xu, S., Ci, Y., Wang, L., Yang, Y., Zhang, L., Xu, C., ... & Shi, L. (2019). Zika virus NS3 is a canonical RNA helicase stimulated by NS5 RNA polymerase. *Nucleic acids research*, *47*(16), 8693-8707.

Yap, L. J., Luo, D., Chung, K. Y., Lim, S. P., Bodenreider, C., Noble, C., ... & Lescar, J. (2010). Crystal structure of the dengue virus methyltransferase bound to a 5′-capped octameric RNA. *PloS one*, *5*(9), e12836.

Yockey, L. J., Varela, L., Rakib, T., Khoury-Hanold, W., Fink, S. L., Stutz, B., ... & Iwasaki, A. (2016). Vaginal exposure to Zika virus during pregnancy leads to fetal brain infection. *Cell*, *166*(5), 1247-1256.

Yon, C., Teramoto, T., Mueller, N., Phelan, J., Ganesh, V. K., Murthy, K. H., & Padmanabhan, R. (2005). Modulation of the nucleoside triphosphatase/RNA helicase and 5′-RNA triphosphatase activities of Dengue virus type 2 nonstructural protein 3 (NS3) by interaction with NS5, the RNA-dependent RNA polymerase. *Journal of Biological Chemistry*, *280*(29), 27412-27419.

Yu, L., & Markoff, L. (2005). The topology of bulges in the long stem of the flavivirus 3′ stem-loop is a major determinant of RNA replication competence. *Journal of virology*, *79*(4), 2309-2324.

Yu, L., Nomaguchi, M., Padmanabhan, R., & Markoff, L. (2008). Specific requirements for elements of the 5′ and 3′ terminal regions in flavivirus RNA synthesis and viral replication. *Virology*, *374*(1), 170-185.

Yu, X., Jin, L., & Zhou, Z. H. (2008). 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature*, *453*(7193), 415-419.

Zamoto-Niikura, A., Terasaki, K., Ikegami, T., Peters, C. J., & Makino, S. (2009). Rift valley fever virus L protein forms a biologically active oligomer. *Journal of virology*, *83*(24), 12779-12789.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of structural biology*, *193*(1), 1-12.

Zhang, K., Li, M., & Sun, F. (2011). Gautomatch: an efficient and convenient gpu-based automatic particle selection program. (Unpublished work)

Zhang, X., Settembre, E., Xu, C., Dormitzer, P. R., Bellamy, R., Harrison, S. C., & Grigorieff, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proceedings of the National Academy of Sciences*, *105*(6), 1867-1872.

Zhang, Y., Corver, J., Chipman, P. R., Zhang, W., Pletnev, S. V., Sedlak, D., ... & Rossmann, M. G. (2003). Structures of immature flavivirus particles. *The EMBO journal*, *22*(11), 2604-2613.

Zhao, B., Yi, G., Du, F., Chuang, Y. C., Vaughan, R. C., Sankaran, B., ... & Li, P. (2017). Structure and function of the Zika virus full-length NS5 protein. *Nature communications*, *8*(1), 1-9.

Zhao, Y., Soh, T. S., Zheng, J., Chan, K. W. K., Phoo, W. W., Lee, C. C., ... & Shi, P. Y. (2015). A crystal structure of the dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog*, *11*(3), e1004682.

Zhao, Y., Soh, T. S., Zheng, J., Chan, K. W. K., Phoo, W. W., Lee, C. C., ... & Shi, P. Y. (2015). A crystal structure of the dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog*, *11*(3), e1004682.

Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., ... & Li, H. (2007). Structure and function of flavivirus NS5 methyltransferase. *Journal of virology*, *81*(8), 3891-3903.

Ziv, O., Gabryelska, M. M., Lun, A. T., Gebert, L. F., Sheu-Gruttadauria, J., Meredith, L. W., ... & Goodfellow, I. (2018). COMRADES determines in vivo RNA structures and interactions. *Nature methods*, *15*(10), 785-788.

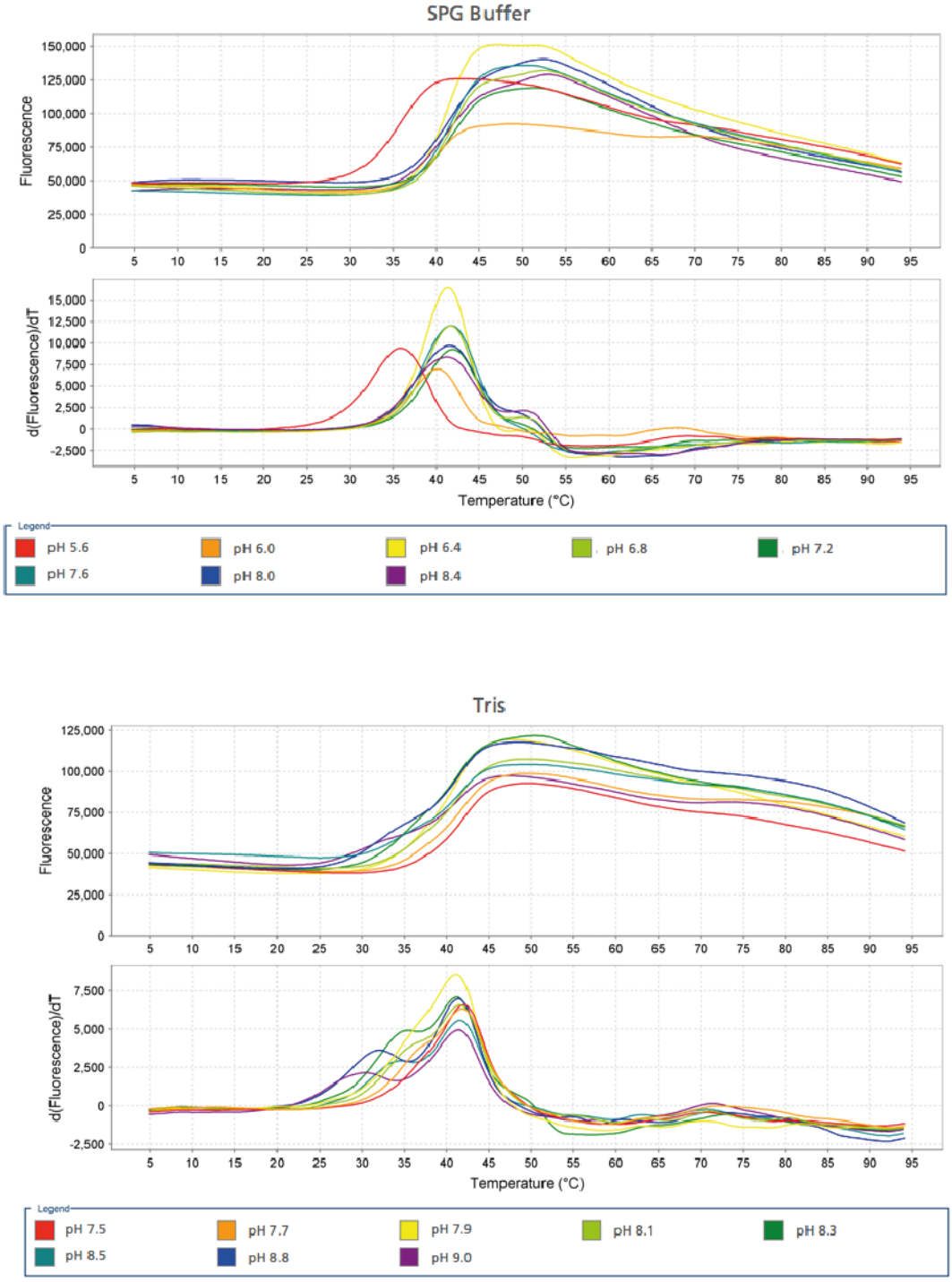
Zmurko, J., Marques, R. E., Schols, D., Verbeken, E., Kaptein, S. J., & Neyts, J. (2016). The viral polymerase inhibitor 7-deaza-2'-C-methyladenosine is a potent inhibitor of in vitro Zika virus replication and delays disease progression in a robust mouse infection model. *PLoS neglected tropical diseases*, *10*(5), e0004695.

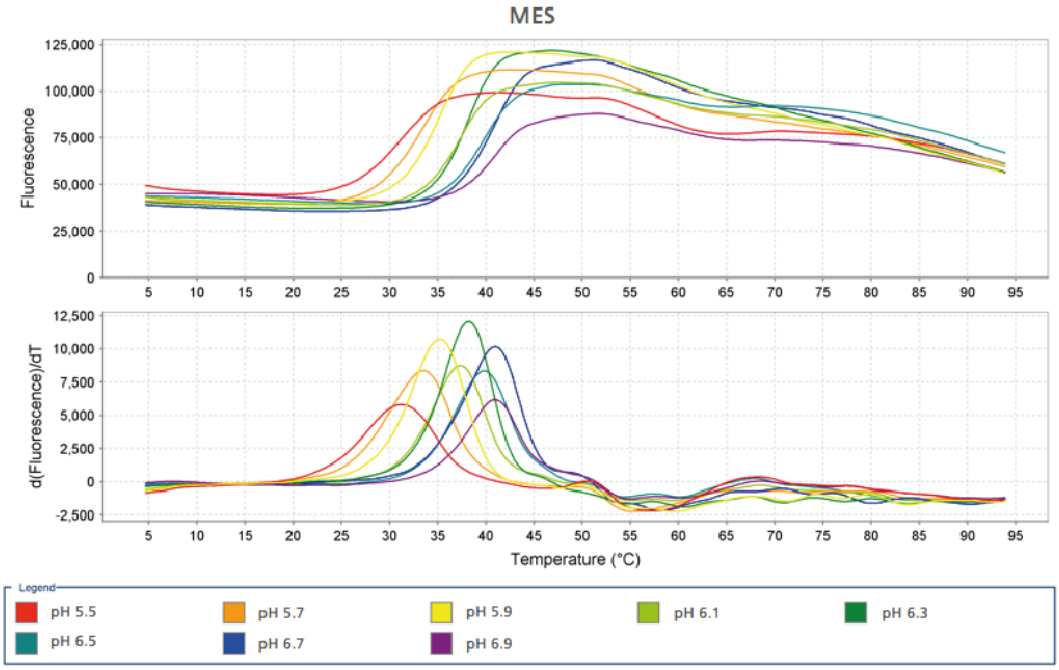
Zonneveld, R., Roosblad, J., van Staveren, J. W., Wilschut, J. C., Vreden, S. G., & Codrington, J. (2016). Three atypical lethal cases associated with acute Zika virus infection in Suriname. *IDCases*, *5*, 49-53.

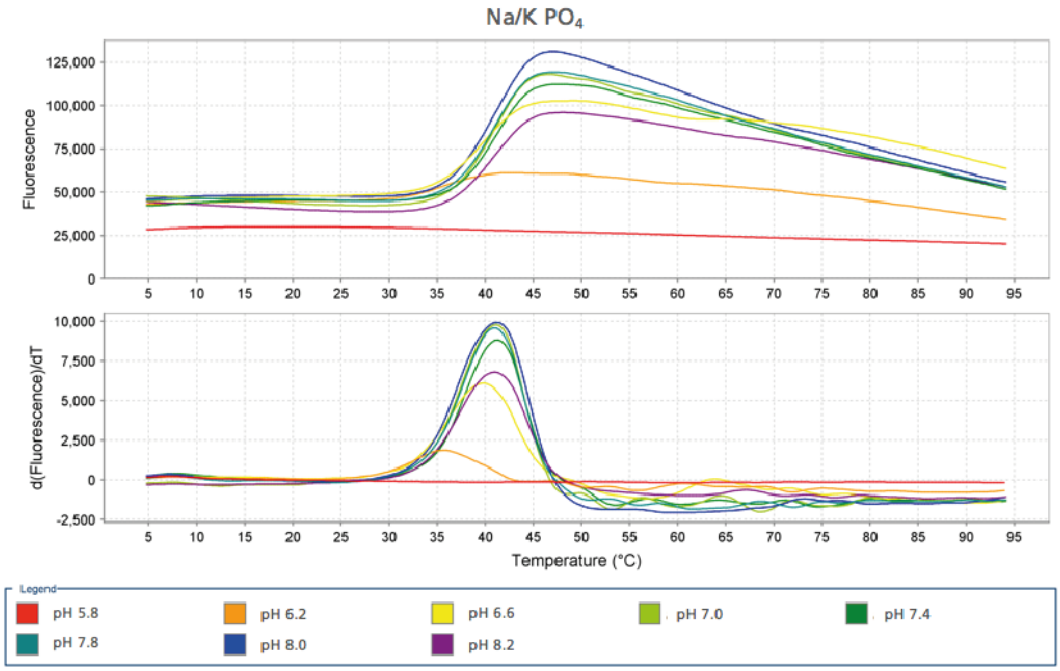
Zou, G., Chen, Y. L., Dong, H., Lim, C. C., Yap, L. J., Yau, Y. H., ... & Shi, P. Y. (2011). Functional analysis of two cavities in flavivirus NS5 polymerase. *Journal of Biological Chemistry*, *286*(16), 14362-14372.

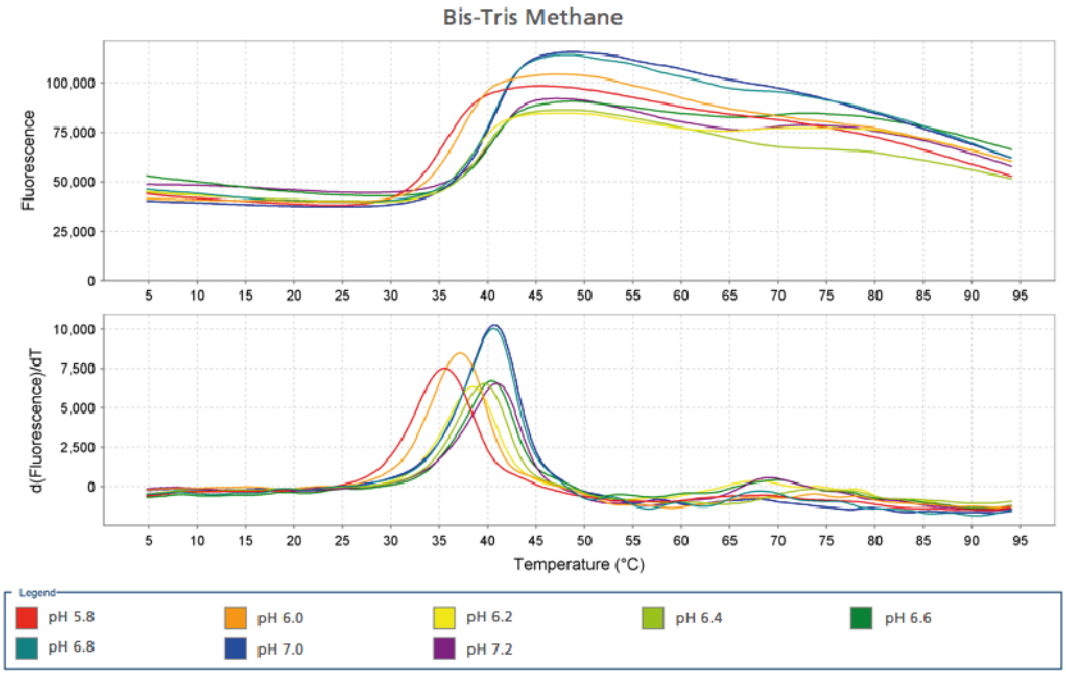
8 Annex

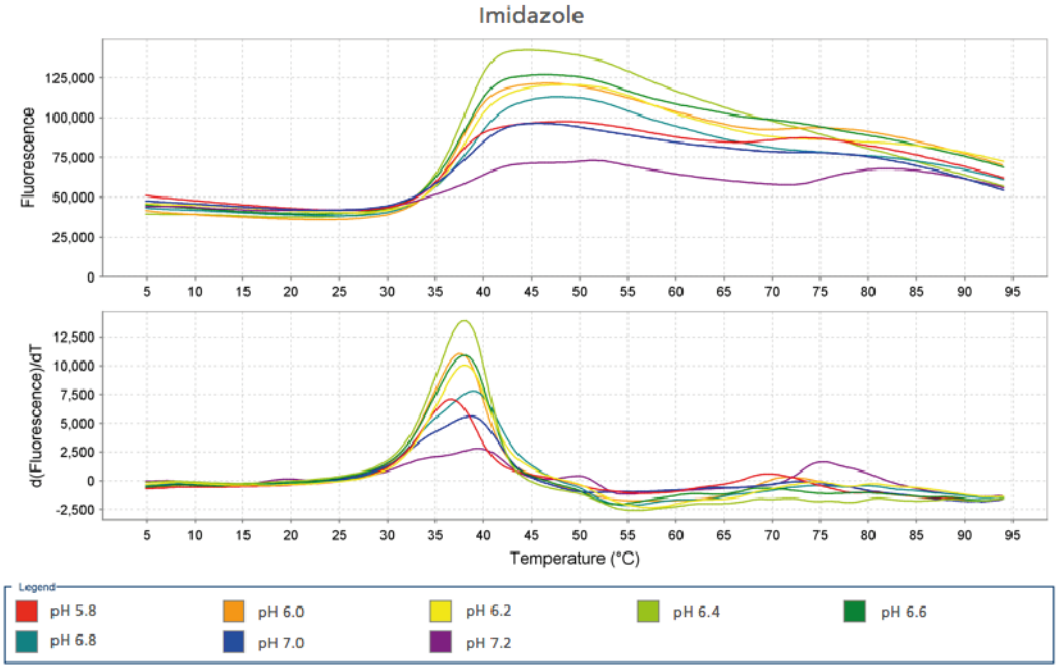
Figure annex 1: Cited on: **3.2.4 Optimization of NS5-5SLA sample for cryo-EM**. Differential Scanning Fluorimetry (DSF) of NS5-5SLA in different buffer conditions. Top panels correspond to the fluorescence increment while bottom panels correspond to the first derivative of the fluorescence.

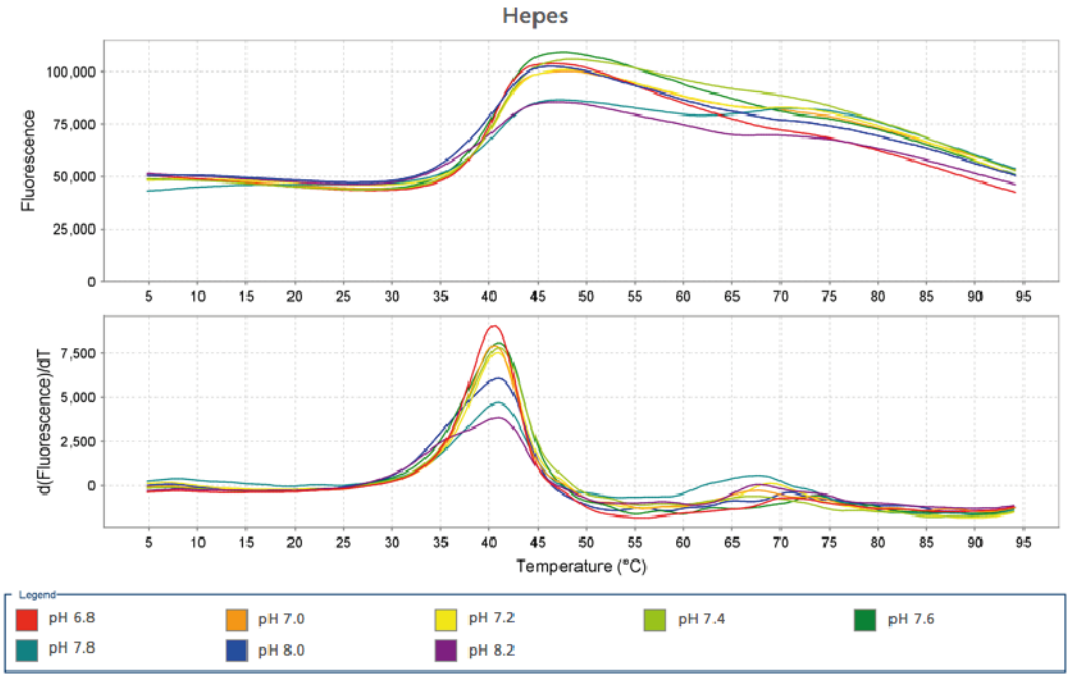



MES
Fluorescence
125,000
100,000
75,000
50,000
25,000
0
d(Fluorescence)/dT
12,500
10,000
7,500
5,000
2,500
0
-2,500
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
Legend
pH 5.5
pH 5.7
pH 5.9
pH 6.1
pH 6.3
pH 6.5
pH 6.7
pH 6.9


Na/K PO4
Fluorescence
125,000
100,000
75,000
50,000
25,000
0
d(Fluorescence)/dT
10,000
7,500
5,000
2,500
0
-2,500
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
Legend
pH 5.8
pH 6.2
pH 6.6
pH 7.0
pH 7.4
pH 7.8
pH 8.0
pH 8.2


Bis-Tris Methane
Fluorescence
100,000
75,000
50,000
25,000
0
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
d(Fluorescence)/dT
10,000
7,500
5,000
2,500
0
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
Legend
pH 5.8
pH 6.0
pH 6.2
pH 6.4
pH 6.6
pH 6.8
pH 7.0
pH 7.2


Imidazole
Fluorescence
125,000
100,000
75,000
50,000
25,000
0
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
d(Fluorescence)/dT
12,500
10,000
7,500
5,000
2,500
0
-2,500
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
Legend
pH 5.8
pH 6.0
pH 6.2
pH 6.4
pH 6.6
pH 6.8
pH 7.0
pH 7.2


Hepes
Fluorescence
100,000
75,000
50,000
25,000
0
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
d(Fluorescence)/dT
7,500
5,000
2,500
0
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
Legend
pH 6.8
pH 7.0
pH 7.2
pH 7.4
pH 7.6
pH 7.8
pH 8.0
pH 8.2

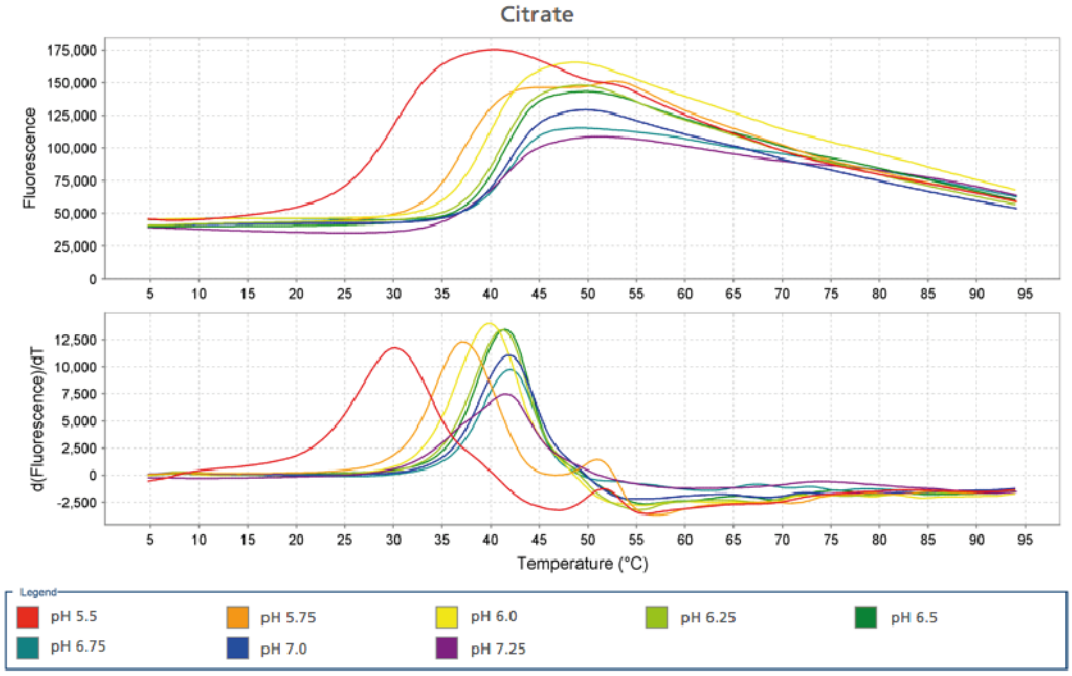

Citrate
Fluorescence
175,000
150,000
125,000
100,000
75,000
50,000
25,000
0
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
d(Fluorescence)/dT
12,500
10,000
7,500
5,000
2,500
0
-2,500
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
Legend
pH 5.5
pH 5.75
pH 6.0
pH 6.25
pH 6.5
pH 6.75
pH 7.0
pH 7.25

Figure annex 2: Cited on: **Biochemical validation of NS5-5SLA interfaces**. Differential scanning fluorimetry (DSF) data of NS5 5SLA binding mutants. Curves represent the mean of the direct fluorescence readings of three replicates (top panel) and the first derivative of the fluorescence (bottom panel). Melting temperatures were: 44 ºC for WT, K226E/K231E and K843E/R844D and 40 ºC for Y840A/R856D, Y840A/K843E/R844D/R856D and K226E/K231E/Y840A/K843E/R844D/R856D.